

**A User's Guide to
FLEXIFIT**

**A model-free analysis technique for families
of curves of similar shape**

**Peter J. Munson, Kim T. Chen
Vincenzo Guardabasso, David Rodbard**

**Analytical Biostatistics Section
Division of Computer Research and Technology
National Institutes of Health
Bethesda, Maryland 20892 USA**

**July 3, 1991
Reprinted: January 1993**

**A User's Guide to
FLEXIFIT**

**A model-free analysis technique for families
of similar shape**

**Peter J. Munson, Kim T. Chen
Vincenzo Guardabasso, David Rodbard**

Analytical Biostatistics Section
Division of Computer Research and Technology
National Institutes of Health
Bethesda, Maryland 20892, USA

Program version 1.0
First Printing Date: November 1990
Reprinted: July 1991
Reprinted: January 1993

Table of Contents

Program Availability	iii
Flexifit License Agreement	iv
1 Introduction	1
2 Required Hardware	5
IBM-PC	5
Apple Macintosh	5
3 Data Files	7
Input Data Format	7
Files Created by Flexifit	8
4 Macintosh Operation	9
Flexifit Menus	9
The Apple Menu	9
The File Menu	9
The Edit Menu	10
The Windows Menu	11
The Flexifit Menu	12
Flexifit Windows	12

The Flexifit Screen Output Window	13
The Flexifit Parameter Values Window	13
The Data Points Window	13
The Graph Window	13
The Reduced View Window	13
Options	14
Setting Parameter Constraints	16
Setting Shared Constraints	16
Setting Constant Constraints	16
Setting Parameter Values	17
Creating Plots	17
Logistic Curves Example	20
5 IBM-PC Operation	29
Program Options ("Switches")	29
Logistic Curves Example	30
Generalized Logistic Example	37
Creating Plots on the IBM-PC	43
Change Command	44
Next Command	45
First Command	45
Key Command	45
Print Command	46
Creating a PostScript File	47
Dot Matrix Printing	47
Creating Lotus 123 Format Files	47
Restart Command	47
Quit Command	47
Prt-Screen Keyboard Key	47
GRAFIT file format	48
Index	51
Appendix A Error Messages	53
Appendix B Reprints	55

1 Introduction

FLEXIFIT is a computer program designed for analysis of families of curves of similar shape. It allows one to abstract the underlying shape of the curves and then characterize the

differences between family members as horizontal and vertical translations and stretches (shifts and scales). Each member of a family is characterized in relation to its first member or to a "template curve".

The common "shape" is characterized as a cubic smoothing spline, applied to all of the data curves after they have been superimposed by appropriate shifting and scaling. Once the shape is determined, a new set of shifting and scaling parameters can be determined. Thus, the algorithm is iterative in nature. Initial guesses of appropriate shifts and scales are required, but these are usually easily obtained by inspection of a graph of the data. Initial estimate rules are built into the program.

The degree of smoothing for the spline may be determined automatically by the program or specified by the user. If the user specifies "monotonicity", the program finds the "roughest" curve consistent with this requirement. The user may also specify the number of inflection points in the final curve, and again, the program will find the "roughest" curve consistent with this requirement. (This latter option may have problems in some cases, though). Alternately, the user may specify the value of the penalty factor l in the penalized sum of squares which is minimized by the cubic smoothing spline. The expression is:

$$(y - o(y,^))2 + l f'2$$

Large values of l make the spline nearly linear. This would be equivalent to fitting the family of curves with a series of straight lines. Small values of l require the spline to nearly interpolate each data point in the family. Extremely small l values make the estimation of the shift and scale parameters impossible, as an interpolating spline may be drawn through the superimposed curves no matter how the data were first shifted and scaled. As a result, one should choose intermediate l values.

To help in the appropriate choice of l , a parameter called "Percentage smoothing" is calculated. It is the ratio of sum-of-squares for a particular l to sum-of-squares for l infinite. This ratio is a measure of the degree of smoothing applied to the data. By this measure, a straight line is 100% smooth.

The equations to be fit to the family of curves may be written:

$$\begin{aligned}
y &= (A1-D1) fs(B1(z-C1)) + D1 \\
y &= (A2-D2) fs(B2(z-C2)) + D2 \\
y &= (A3-D3) fs(B3(z-C3)) + D3 \\
&\cdot \\
&\cdot
\end{aligned}$$

where fs is the common cubic spline function; the As and Ds are "vertical" parameters, and the Bs and Cs are "horizontal" scaling parameters. This parameterization was inspired by the "four-parameter logistic" used in our ALLFIT program, namely,

$$y = f(A - D, 1 + (x/C)^B) + D,$$

which can also be written,

$$y = (A-D) fl(B(z - \log C)) + D$$

where $z = \log(x)$ and the logistic function fl is defined

$$fl(z) = 1/(1+10^z)$$

Notice a subtle difference between the distinct uses of the letter "C", namely between C in the FLEXIFIT equation and logC in the ALLFIT equation. This confusion is unfortunate, but now exists in the published literature. Therefore, we shall refer to the FLEXIFIT C (CFLEXIFIT) and to the ALLFIT C (CALLFIT). The following equivalence holds:

$$CFLEXIFIT = \text{Log}_{10} (CALLFIT)$$

$$CALLFIT = 10CFLEXIFIT$$

If you are doing problems related to biopotency estimation, ED50, IC50, or LD50 analysis, analysis of pharmacological dose-response curves, binding inhibition curves, bioassay, radioimmunoassay, ELISA, IRMA, EMIT, USERIA, or similar procedure, in short, if you might have used ALLFIT, then use the ALLFIT parameterization. CALLFIT can usually be interpreted as the ED50 in the original data scale.

If you are applying FLEXIFIT in other contexts, where you seek to

graphically shift and scale the curves so that they superimpose, the CFLEXIFIT will correspond to the horizontal shift parameter, B to the horizontal scaling parameter.

In any case, think of D as a vertical shift parameter (e.g. a baseline value) and A as a maximal value. The difference (A-D) itself is a vertical scale parameter, or may be thought of as the vertical excursion or data range. A and D are most meaningful when the shape is like that of a logistic function, i.e. having both a lower and an upper plateau. For descending logistic-like curves, A is the level of the low dose (small z) plateau, and D is the level of the high dose (large z) plateau (provided you have used the appropriate Template or Reference curve, see below).

The FLEXIFIT problem is initially overparameterized (for example, think of a family with only one member; then A1,B1,C1 and D1 are undefined since a spline can be fit regardless of their values!) We remove these excess parameters by establishing one curve as a basis of reference for all the others. This usually is accomplished by having A1, B1, C1 and D1 set to constant values. It is often convenient to set A=1, D=0, B=1, C=0 for this reference curve.

One of the features of FLEXIFIT allows one to constrain various parameters to share values. It is possible, for example, to require the first three curves to share a baseline (D) value, while allowing the rest of the curves to have individual baseline levels. One may also remove parameters from the model by fixing them at constant values. Careful use of the parameter sharing and fixing facility allows for the comparison of a wide variety of statistical hypotheses (e.g. do compound I and compound II have the same ED50 value).

In the case of sigmoidal or logistic-like curves, the optional logistic

$$T(z) = f(10z, 1+10z)$$

should be applied before calculating the spline. Thus, the model now becomes

$$y = (A-D) fs(T(B(z-C))) + D$$

or $y = (A-D) fs(T(z')) + D$

where $z' = B(z-C)$

Now, the shifting and scaling is done in one scale (the $z = \log(x)$ scale) while the spline is fit in another (y vs. $T(z')$).

To the user, the use of a transformation is transparent, except that more "realistic" plateaus are fit with the above transformation. Moreover, using the above template curve together with this transformation and specifying $PV=1000$, results exactly comparable to those from ALLFIT are obtained. In this sense, FLEXIFIT becomes a true extension of the four-parameter logistic model.

Selecting the logistic transform option of the program automatically generates a special three-point template curve with corresponding parameters $A0$, $B0$, $C0$, and $D0$. These parameters are always set to constant values. The purpose of this curve is to provide three artificial data points to "anchor" the spline at three places: the beginning, middle, and end. Effectively, the spline, in the y vs. T coordinate system is constrained to pass through $(0,1)$, $(.5,.5)$ and $(1,0)$.

A second transformation is provided, which simply takes the base 10 logarithm of any input "x" data. This is used primarily to make FLEXIFIT directly accept data files prepared for the ALLFIT program. Said another way, if the user of FLEXIFIT desires to make shifts and scale changes in the logarithm of the original data, as is the case when x represents concentration, he may do so with this option.

2 Required Hardware

IBM-PC

The FLEXIFIT program is designed to run on all IBM-PC and related clones. The program has been run on an IBM PS2/80(80386), Compaq Plus (8086), Compaq III (80286), and IBM XT computers. It failed

for unknown reasons on an IBM XT with Tiny Turbo 80286 card. We do know that using different compiler options (specifically, compilation using 80286 codes) does work on the Tiny Turbo card.

The program is designed to make optional use of the 8087 (or appropriate) floating point coprocessor chip, and installation of this chip is recommended as it greatly speeds up processing.

Execution times for the examples shown in this manual run from nearly instantaneous (80386/80387) to 15 minutes (8086 w/o 8087).

Use of automatic selection of smoothing parameter can increase run times up to 15 fold.

A full 512K or 640K machine is recommended for program operation, as is a hard disk, video board (CGA or Hercules) for the graphics program. A dot-matrix or laser printer is useful to record the graphs onto paper. The utility program GRAFTRAX.EXE is provided to dump CGA memory to an Epsom FX80) or equivalent printer.

Apple Macintosh

Flexifit is also available for the Apple Macintosh. It has been tested on the Mac II, the Mac IICx and the Mac SE. A minimum of 1 meg. of memory is suggested. Processing speeds on the Mac II are comparable to those seen on the IBM-AT. No installation is required on the Macintosh. The FLEXIFIT program can be run from floppy or copied onto and launched from the hard disk. It will run faster on the hard disk because file I/O is faster on the hard disk than on the floppy. As with any program, it is suggested that you make a backup copy of FLEXIFIT (either onto hard disk or floppy) and run the program from your backup.

3 Data Files

Input Data Format

The format of the input data file is the same for the PC and the

Macintosh versions of Flexifit. The data files used as input to FLEXIFIT are text (ASCII) files which can be created using any text editor. X,Y pairs are placed on each line where X is the independent value (abscissa), usually in terms of concentration if you are working with a bioassay or dose-response study, and Y is the dependent (ordinate) value, or response of the system in bioassay. The particular interpretation, and units of X and Y are, of course dependent on your particular study.

Each data file may contain a number of separate curves, or you may put a single curve in each data file. Curves within data files start with a line of text (don't use numerals in the first position as this will be interpreted as a continuation of a previous curve). It is smart to use a short, meaningful description on this line for each curve.

The individual (X,Y) pairs are placed on separate lines. Each number may be separated by blanks, tabs or commas. No particular format or position within the line is required. The sample data set used in this document is listed below.

Data file sample.all

APOMORPHINE-TER***

0, 12570
1E-9, 12208
1E-8, 11789
3E-8, 11273
1E-7, 10382
3E-7, 9828
1E-6, 7918
3E-6, 6351
1E-5, 6135
1E-4, 5724

DOPAMINE-TER***

0, 12556
1E-9, 12421
1E-8, 11743
3E-8, 11287
1E-7, 11333
3E-7, 10328
1E-6, 9443
3E-6, 8610

1E-5, 7853
1E-4, 5984
EPINEPHRINE-TER***
0, 12940
1E-8, 12090
1E-7, 11248
3E-7, 10628
1E-6, 10239
3E-6, 9282
1E-5, 8526
3E-5, 7560
1E-4, 6788
1E-3, 6140
NOREPINEPHINE-TER***
0, 12923
1E-8, 12421
1E-7, 11625
3E-7, 10955
1E-6, 10196
3E-6, 9799
1E-5, 8905
3E-5, 8202
1E-4, 7289
1E-3, 6416
Files Created by Flexifit

Flexifit produces two files; a session file and a graph file. (The graph file is not produced by the Macintosh version).

The session file is a text file which contains the textual results output to the screen (the Flexifit Display Screen window for the Macintosh version) by the Flexifit program. This file contains the switches or options, parameter values and constraints and the results of the fit.

The graph file is only produced by the PC version. It is later used as input to the GRAFIT program on the PC. The format of this file is discussed in the section of PC Operations.

4 Macintosh Operation

A typical FLEXIFIT session on the Macintosh would involve selecting a data file and performing a fit. The program generated

results are displayed in several different windows on the screen. When the fit is complete you can plot the results by selecting Plot from the Flexifit menu. The graphs are displayed in a graph window. The total plotting area is greater than that displayed in the graph window. Another window, Reduced view, can be used to look at the entire plotting area. A more advanced user may want to change an option and/or set shared or constant parameters before performing the fit. All of the menus and windows are discussed in the sections that follow. This section on Macintosh Operations ends with a sample session.

Flexifit Menus

The Flexifit program uses five menus; the Apple Menu, a File Menu, an Edit Menu, a Windows Menu and lastly the Flexifit Menu. These menus are discussed below.

The Apple Menu

The Apple menu gives you access to the About the Finder window and any desk accessories installed on your Macintosh. (In addition, with MultiFinder on, you can use the Apple menu to move among open applications. For details, see chapter 10 of the Macintosh System Software User's Guide.)

The File Menu

Select File. Select File causes the standard file dialog box to be displayed. Only files with the .ALL file extender are displayed. Double click the file to open it. (Alternately you can select the file and then click the Open button). The data from the file is displayed in the Data Points window as it is read.

Save As... . Save As allows you to save information displayed on the screen into a file. Plots are saved as .pict files. Text from any of the editors is saved as text files.

Print... . Print allows you to print text from the editors ; and plots from the graph window.

Save Settings... . Save Settings is used to save the placement of windows on the screen.

Quit Flexifit. Quit Flexifit closes the session file and returns you to the finder.

The Edit Menu

Cut. Cut removes the selected material and places it on the Clipboard (replacing the previous contents, if any).

Copy. Copy puts a copy of the selected material on the Clipboard (replacing the previous contents, if any). The "original" stays where it was when you selected it.

Paste. Paste puts a copy of the contents of the Clipboard at the insertion point. You can continue to paste copies until you cut or copy a new selection (which replaces the old contents of the Clipboard).

Clear. Clear removes the selected material without placing it on the Clipboard. The contents of the Clipboard remain intact.

Select All. This command selects all material in the active window. All plots in the plot window or all text in the Flexifit Screen Display.

Combine. Combine is used to merge pictures into a single picture. Each plot is drawn as a new picture and any single plot can be saved in a .pict file. If you wanted to save more than one plot in a .pict file then you would select the plots (click and then shift click) and then use combine to merge the pictures. These plots could then be saved in a single .pict file using the save command.

The Windows Menu

Graph Setup... Graph Setup displays a dialog used to set defaults used in the graph window. This menu item is displayed when the graph window is active. If the Flexifit Screen Display is the active window then Edit Setup... will be displayed instead in this menu position. Sheet Setup... is displayed if the Parameter Values Spread Sheet is the active window.

Clipboard. Clipboard can be used to find out how much memory you have left and also shows you what text or picture the clipboard contains.

Reduced View. Reduced View displays a window showing the entire drawing area.

Graph .Graph displays the graph window which is used for plotting. Plots can not be generated until after a fit has been completed successfully.

Data Points. Data Points displays the window where the input data is displayed.

Flexifit Screen Output. Flexifit Screen Output displays the window where the tabular results of the fit are displayed.

The Flexifit Menu

Options. Clicking the Options menu item results in the display of the options dialog box.

Constraints. The constraints menu selection is used to set and clear constant and shared parameters. Clicking the constraints menu item results in the display of a sub menu with the choices Constant, Shared and Clear. Selecting Constant or Shared results in the display of a dialog box. Clear removes all constant and shared associations but does not recalculate the parameter values.

Parameter Values. Clicking the Parameter Values menu selection results in the display of a spread sheet containing the parameter values. The values may be edited using this spread sheet. If the value of a shared parameter is changed the change is automatically applied to the associated shared parameters.

Reset Parameters. Reset Parameters initializes the values of all parameter values. Shared and constant associations are cleared.

Fit Data. Fit Data begins the data fit. Results are displayed on the monitor as they are obtained. Parameter values as displayed in the spread sheet are updated as the fit progresses. (Once the fit has started it may be interrupted by holding down the command key and striking the letter 'c').

Plot. Plot results in the display of a submenu containing the three types of plots... Fitted Curves, Common Spline - Linearized and Common Spline. Upon selecting a plot from the plot submenu, a dialog box is displayed that allows you to customize the look of the plot.

Plot input. Plot input allows you to plot the point in the original data file. This can be done before any fit has been made.

Flexifit Windows

The Flexifit program uses five window.

- Flexifit Screen Output
- Flexifit Parameter Values
- Data Points
- Graph
- Reduced View

You can choose to display or not display any of these windows. The windows are removed from the screen by clicking the goaway box located in the upper left hand corner of the window. To redisplay the window click on the appropriate selection in the Windows Menu. All of the windows can be resized except for the reduced view window. Upon launching the Flexifit program you may want only certain windows to be displayed at a particular size and location on the screen. This can be done by first arranging the windows and then clicking Save settings... from the File Menu. The windows are discussed below in greater detail.

The Flexifit Screen Output Window

The Flexifit Screen Output Window is used to display all information pertinent to a fit. This information includes the name of the input file, the number of curves and data points in the input, the options, constraints and parameter values and finally the results of the fit. This data is also saved automatically in a session file.

The Flexifit Parameter Values Window

The Flexifit Parameter Values Window functions like a spreadsheet. You may edit parameter values by clicking on the value, entering the new value and then hitting return. If the parameter value changed is a shared parameter, all other parameters in the same group will be updated to reflect the new value. As the parameter values change during the course of the fit, the new values are updated in the spreadsheet.

The Data Points Window

The Flexifit Data Points Window is used to display the data points from the input data file. This window is used for display only and can not be used to edit the data file.

The Graph Window

The Flexifit Graph Window is used to plot the results of the fit. Only a portion of the entire drawing area is displayed in this window. The entire drawing area is broken up into pages. Both the size and number of pages can be set via the "Graph Setup..." dialog obtained by selecting Graph Setup... in the windows menu. Different plots can be drawn on different pages. It is easy to move from page to page via the Tab key. Scroll bars can be activated by pressing and holding the option key. The option key also activates a "hand cursor" which is used to drag the drawing area across the screen. If you use the default plot settings, then each plot will be drawn on a separate page. You may change this by resizing the plot box (click and drag) relocating it and then unchecking 'Use default frame' in the plot setup dialog. The plot will then be resized and drawn at the location you specified.

The Reduced View Window

The Reduced View Window is used to display the entire drawing area and the pages into which the drawing area has been broken. This is useful to look at all the plots you have created and see how they will be arranged on a page(s) if and when they are printed. The numbers displayed at the top of the window, 1:6 for example, is a ratio of the reduced view window to the graph window. A "magnifying glass" cursor (option key) when clicked in the reduced

view, causes the corresponding graph drawing area to become visible in the graph window.

Options

FLEXIFIT options are changed via the dialog box shown below. Click the menu item Options to display this dialog.

Is monotonicity required ?

Checking this box requires the curve to be monotonic (also controls the degree of smoothing)

Specify maximum inflection points ?

The number of inflection points is used to control smoothing.

Specify minimum % smoothing ?

Provides a minimum degree of smoothing. 100% smoothing produces a straight line.

Fit bell shaped curve

This is a special feature for curves which seem to rise, then fall back to the original value.

Specify starting Lambda ?

Lambda also controls the degree of smoothing. Values of lambda are tried iteratively until the other smoothing criteria are met. Starting values of lambda should be large, say $1E3$.

Specify ending Lambda ?

Final or ending values should be small, around $1E-3$.

Specify fixed Lambda ?

Specifying a fixed value for lambda reduces computing time considerably.

Create session file ?

- Automatically writes the tabular results of the session to

a file. The file names are FLEXnn.SES where nn is incremented for each run of FLEXIFIT.

Apply template ?

A template curve should be generated whenever the logistic transform is used. The purpose of the template is to "nail down" the ends of the spline in the transformed coordinate system. Without a template, poor convergence may be experienced.

Take logs of data ?

Take logs of X values in input file.

Use logistic transform ?

Useful if data represent a logistic (S-shaped) curve characteristic of dose-response curves plotted on log X scale.

Weighted fit ?

Allows for user-specified variance model to generate weights for regression. The variance model assumes a quadratic relationship between Y and its variance: $\text{Var}(Y) = A_0 + A_1Y + A_2Y^2$

Setting Parameter Constraints

A set of four parameters (A, B, C and D) is associated with each curve in a data set. Constant and shared constraints can be associated with these parameters. A data file must be selected before constraints can be set. After selecting a data file the Constraints menu item in the Flexifit menu will be ungreyed. Shared and constant constraints can then be set by clicking the Shared or Constant menu items displayed in the Constraints sub menu. Setting the shared and constant constraints is discussed below.

Setting Shared Constraints

First click the Constraints menu item and then click the Shared menu item from the sub menu. The dialog box shown below will be displayed. (The dialog box shown is for a data set with four data curves).

A1 can only be shared with other A parameters; B1 with other B parameters etc. Thus A1 cannot be shared with B1, C1 or D1. To set A1 and A2 as shared simply click the boxes and then click Enter or Done button. Click Enter if you want to select more shared groups. Clicking the Done button will remove the dialog box from the screen. Once a group of parameters is selected as a group, they will be greyed. Clicking the Cancel button will undo the shared associations you just made.

In order to clear any or all shared constraints you must select Reset parameters from the Flexifit Menu.

Setting Constant Constraints

To set parameter values to constants Click the Constant menu item from the Constraints sub menu. Parameters are set to constants by clicking the appropriate check box in the Constant Parameters Dialog. The constant constraint can be cleared by unchecking the check box.

Setting Parameter Values

Parameter values can be modified in the parameter values spread sheet. If the parameter value changed is a shared parameter then its associated shared parameters will be automatically updated. The parameter values spread sheet is displayed by clicking the Flexifit Parameter Values menu item in the Windows menu or by clicking the Parameter Values menu item from the Flexifit menu.

Creating Plots

FLEXIFIT offers three types of plots: Fitted Curves, Common Spline and Common Spline - Linearized. To create a plot, click on the desired type of plot from the plot submenu. The following dialog box will be displayed:

You can use the default settings. Simply click the OK button and the plot will be drawn in the graph window.

Changing Labels on Plots

The buttons ALL, Title etc. located at the bottom of the dialog box are used to change the font, color, size, and textual content of the plot labels. Upon clicking any of these buttons you will be presented with a dialog similar to the one below:

Changing the Size of the Plot

The size of the frame (plot box) in which the plot is drawn can be altered by click/ dragging the frame prior to selecting the plot. You must then clear the check box labeled Use default frame (in the plot dialog box) else the default size and location will be used.

Plots are drawn in a full graph window and a reduced view window. The graph window displays only a portion of the total drawing area. The reduced view window (Selected from the Window menu) displays the entire drawing area, broken up into one or more pages. The Fitted Curves Plot is drawn on page 1, the Common Spline on page 2 and the Common Spline - Linearized is drawn on page 3. When these are printed they will be printed on separate pages. You can set the number and size of each page via the Graph Setup Dialog. (To obtain this dialog bring the graph window to the front and click Graph Setup under the Window menu). You can move from page to page via the Tab key, scroll bars (press option key), a "hand" cursor which drags the drawing area across the graph window, and a "magnifying glass" which, when clicked in the

reduced view, causes the corresponding graph drawing area to become visible in the full view.

Printing Plots

To print a plot, first make sure that the graph or reduced view is the front window and then select print from the File menu.

The Macintosh screen resolution of 72 pixels/inch is lower than what most printers can achieve. To work at the higher printing resolutions (more pixels per inch), you can choose a smaller "percent reduction". Choosing 50% reduction in the "page Setup" dialog, for example, would produce a graph page size which was twice as wide and twice as tall as the printer's normal page size (corresponding to 100%). When printing, these large pages are then reduced to a normal page size without the loss of information (each pixel on the screen appears as a separate dot on the printed page). This technique is particularly useful when used with the Laser Writer, where a 25% reduction produces a graph page size which is large enough to represent every pixel which can be drawn on a LaserWriter page (300 dpi). A graph drawn into such a large page, for example, would be reduced 4-fold in each dimension when printed, resulting in an extremely fine and detailed drawing.

Modifying Plots Using other Software

Once created plots can be combined and/or saved as .pict files which then can later be modified using drawing programs such as MacDraw.

Logistic Curves Example

We now describe a session of FLEXIFIT using the data file SAMPLE from a previous section. The data represent inhibition of binding of a radioactively labeled drug by four unlabeled drugs. This identical data set is described in the ALLFIT User's Guide (November 1988). This run of FLEXIFIT is arranged so that identical results are obtained. Thus, this first approach is a good starting place for previous ALLFIT users. Not described here, but in common with ALLFIT, FLEXIFIT has the potential to constrain the parameters to constant values or to values taken by other parameters. Thus, since the amount of inhibition in the presence of a zero dose of drug is theoretically identical, it might be appropriate to constrain $A_1 = A_2 = A_3 = A_4$. One can also attempt the constraint $D_1 = D_2 = D_3 = D_4$ to test if the

maximal inhibitions for all four drugs are the same. (Note: This example is also described in the PC-Operations chapter.)

We run FLEXIFIT using the logistic transform. Click on the Options selection in the FLEXIFIT menu. The options dialog displayed below will be displayed. Click the Use logistic transform box and then click OK.

Next we select the file. Click on Select file from the File menu. Double click on the file sample.all displayed in the file selection dialog. As the file is read, the data points will be displayed in the Data Points window. The output shown below will be displayed in the Flexifit Screen Display window.

----- FLEXIFIT v. 1.0 ----- 7/16/1990 10:24:19

TEMPLATE CURVE:

CURVE: BLTIN3 --- CURVE # 0 3-POINT BUILT-IN TEMPLATE

***** DATA FILE(S) OPENED *****

DATA FILE: SAMPLE.ALL

CURVE: APOMOR --- CURVE # 1 HAS 10 POINTS.

CURVE: DOPAMI --- CURVE # 2 HAS 10 POINTS.

CURVE: EPINEP --- CURVE # 3 HAS 10 POINTS.

CURVE: NOREPI --- CURVE # 4 HAS 10 POINTS.

***** VALUES OF CONSTRAINTS *****

TRANSFORMATION FORMULA : LOGISTIC

TAKE LOG OF X-DATA : YES

TEMPLATE CURVE INCLUDED : BLTIN3

MONOTONICITY NOT REQUIRED

MAXIMUM # OF INFLECTION POINTS : 44

LAMBDA* ADJUSTED BY PROGRAM

PERCENTAGE SMOOTHING SPECIFIED : 100.0

COEFFICIENTS OF VARIANCE FUNCTION :

A0 : 1.00000

A1 : .000000

A2 : .000000

Only one input data file may be fit at a time. The file may contain as many as 10 curves. The file SAMPLE.ALL contains four data curves, each with 10 points. The computer will take logs of the x-data by default, use the logistic transform, and fit a 100% smooth spline, i.e. a straight line. A 3 point template curve (BLTIN3, Curve #0) is automatically generated and the corresponding values for parameters A0, B0, C0 and D0 will be fixed at pre-determined values. We allow for up to 44 inflection points, the maximum possible with 44 points. Finally, by default we are using an unweighted fit. Values for A0, A1, and A2, which determine the variance model can be entered using the Options Dialog. Once the file is read in, initial values of the parameters are calculated and displayed in the Flexifit Parameters Sheet. (These values are also written to the Flexifit Display Screen and the Session Data File when the fit is started.)

In this session we do not make use of the Shared Parameters, or Constant Parameters features. These features work exactly as they do in ALLFIT, and allow a variety of constraints to be added to the parameters. The parameters for Curve #0, the template, are automatically set to constant values. The computer-derived initial estimates of the parameters can be modified by simply clicking on the parameter you wish to change in the Flexifit Parameters sheet. In most cases, automatic estimates will work, but for more difficult problems, careful, manual readjustment of these values is required.

To start the fit select Fit Data from the Flexifit menu.

```

FIT # 1-----

ITR #  LAMBDA*      SSQ      SSQ_LR  % SMOOTHING
LSQ ITR#
  1  1.000000E+07  1.008755E+06  1.008773E+06  100.00
4
  2  1000000.      1.008602E+06  1.008773E+06  99.98
1
  3  3.162278E+06  1.008718E+06  1.008773E+06  99.99
1

```

4	5.623413E+06	1.008742E+06	1.008773E+06	100.00
1				
5	4.216965E+06	1.008732E+06	1.008773E+06	100.00
1				
6	3.651741E+06	1.008725E+06	1.008773E+06	100.00
1				
7	3.398208E+06	1.008722E+06	1.008773E+06	99.99
1				
8	3.522695E+06	1.008724E+06	1.008773E+06	100.00
1				

---- FINAL LAMBDA, SSQ, AND % SMOOTHING ----
 LAMBDA* 3.522695E+06
 LAMBDA 3.038175E+09
 FINAL WEIGHTED SSQ 1.111305E+06
 FINAL SPLINE SSQ 1.008724E+06
 % SMOOTHING 100.00

The fit is now complete, after 8 major iterations. Within each of the major iterations, a variable number of "least-squares" iterations were made. The goal of the iterations was to bring the final percentage smoothing to the specified amount, in this case 100%. The smoothing is controlled within the program by a parameter l^* , varied between 10^{-7} and 10^7 . Actually, there is a scaled version of this parameter, called l , which is actually used in the spline subroutine. At this point summary tables are printed out.

1/25/1989

10:40: 9

 ***** FINAL RESULTS *****

CONDITIONS : MONOTONIC
 1 INFLECTION POINTS
 WEIGHTED SSQ IN ORIGINAL COORDINATES: 1111305.0
 WEIGHTED SSQ IN SPLINE COORDINATES : 1008724.0
 SSQ DUE TO MERGING : 102581.30
 MEAN SQUARE (MS) : 41159.430
 ROOT MEAN SQUARE (RMS) : 202.87790
 NORMALIZED CURVATURE : 2.16680900E-13

---- RESULT OF THE CURRENT FIT ----

```

=====
==
=====
Curve No. of Weighted  MSSD  Residuals No. of
Residuals             + -  Runs
#  Points Sum of Sqr.
-----
-----
0   3  .4693      N/A   2  1  3  +++
1  10  6.0747E+05  GOOD  5  5  5  +-----+-
++
2  10  2.4385E+05  GOOD  3  7  7  -+-----
+-
3  10  1.5046E+05  GOOD  5  5  5  +-----+-
-+
4  10  1.0952E+05  GOOD  6  4  6  -+-----+-
+
Total  43  1.1113E+06  GOOD  21  22  18  +-----
--+++++-----
-----+-
--+-----+-
=====
==
=====

```

The conditions of the final curve are described: it is monotonic with one inflection point, exactly what we would expect of a logistic curve in the log x scale. The residual sum-of-squares is printed. The "spline SSQ" is measured after merging or combining points with nearly the same x-value, after the shifting and scaling operations have been applied. This is necessary since the spline routine requires distinct, non-identical x-values at every data point. The component of the residual sum-of-squares due to this "merging" is printed. Ordinarily, this should be a small component of the overall sum-of-squares. The RMS error is computed by dividing the SSQ by the degrees of freedom, then taking the square-root. The degrees of freedom are taken to be the total number of points less the number of true parameters. In

this case there are 43 data points and 16 parameters, or 27 degrees of freedom. In this version (2.1) of FLEXIFIT, no correction to the degrees of freedom is made for the smoothing applied to the data. With 100% smoothing, there is no need for this correction, but as this figure drops, the true degrees of freedom also drops. Thus, the RMS printed here is often an underestimate.

The summary table includes an approximate MSSD test, to evaluate the randomness of residuals in each curve, the numbers of positive and negative residuals, the numbers of runs of residuals, and the sequence of signs of residuals in each curve. One can evaluate this table visually to check for patterns which would indicate a lack of fit. No apparent patterns are present in this case. Long clusters of + or - signs indicates systematic departure from the fitted curve. With the logistic option, 100% smoothing, used in this example, we would say that the data appear well described by a four-parameter logistic.

Next the final parameters are tabulated.

1/25/1989 10:40: 9

---- PARAMETER VALUES ----

=====

==

=====

Curve Shared Name Groups	Parameter	Value	Standard Error	RELATIVE Error
--------------------------------	-----------	-------	-------------------	-------------------

BLTIN3	A0	7.2160E+05		
APOMOR	A1	1.2327E+04	150.9	1.2%
DOPAMI	A2	1.2610E+04	189.4	1.5%
EPINEP	A3	1.2888E+04	203.2	1.5%
NOREPI	A4	1.2978E+04	202.4	1.5%

BLTIN3	B0	1.000		
APOMOR	B1	-.7255	7.6341E-02	10.5%
DOPAMI	B2	-.3716	5.7438E-02	15.4%
EPINEP	B3	-.3652	4.5034E-02	12.3%
NOREPI	B4	-.3584	4.7306E-02	13.1%

		<----- ALLFIT 'C' ----->		<---- FLEXIFIT
'C' ---->				
			Value	
Std. Error				
BLTIN3	C0	1.000	.0000	
APOMOR	C1	4.3614E-07	7.3051E-08	16.7% -6.360
		7.2742E-02		
DOPAMI	C2	5.5544E-06	4.3884E-06	79.0% -5.255
		.3431		
EPINEP	C3	4.4418E-06	2.0254E-06	45.6% -5.352
		.1980		
NOREPI	C4	6.4479E-06	3.3727E-06	52.3% -5.191
		.2272		

POTENCY RATIO

C2/C1	12.74	10.29	80.8%
C1/C2	7.8521E-02	6.3416E-02	
C3/C1	10.18	4.947	48.6%
C1/C3	9.8189E-02	4.7698E-02	
C4/C1	14.78	8.120	54.9%
C1/C4	6.7640E-02	3.7150E-02	

BLTIN3	D0	.0000		
APOMOR	D1	5437.	234.0	4.3%
DOPAMI	D2	3773.	1191.	31.5%
EPINEP	D3	5072.	535.2	10.5%
NOREPI	D4	5319.	606.1	11.3%

=====
=====
=====
=====
=====

N.B. The Standard Errors printed here are based on the RMS error tabulated previously. In this version of FLEXIFIT, no correction is made for the loss of degrees of freedom due to spline smoothing. Thus, these Standard Errors are underestimated when the percentage smoothing is less than 100%.

Now, the observed and predicted data are tabulated. Also, a graphics data file is created. The first graph shows the original or LogX version of the curves. Parallelism and the degree of horizontal or vertical shift or scale can be evaluated here. Second, a graph of the super-imposed (i.e. shifted and scaled data) together with the common spline is prepared. Use this graph to evaluate the common "shape" of the family of curves. When the logistic transform is used, a third graph is prepared, showing the common, shifted, scaled data and spline, but in the log-x coordinate system. This third graph is useful for visually evaluating the departure of the common shape from a simple logistic curve.

1/25/1989 10:40: 9

---- ORIGINAL DATA ----

```
=====
==
=====
```

---- CURVE #0 TEMPLATE: BLTIN3

DATA	SPLINE	Log X	ORIG. X	Y	Y CALC	RESIDUAL	WEIGHT
#	KNOT	#					
1	1	-6.00	1.000E-06	.000	-.337	.337	1.00
2	16	.000	1.00	3.608E+05	3.608E+05	-.563	1.00
3	38	6.00	1.000E+06	7.216E+05	7.216E+05	.250	1.00

```

---- CURVE # 1: APOMOR
DATA SPLINE Log X ORIG. X Y Y CALC
RESIDUAL WEIGHT
# KNOT #
-----
-----
 4 38 Neg. Inf. .000 1.257E+04 1.233E+04 243.
1.00
 5 37 -9.00 1.000E-09 1.221E+04 1.224E+04 -35.8
1.00
 6 35 -8.00 1.000E-08 1.179E+04 1.191E+04 -119.
1.00
 7 31 -7.52 3.000E-08 1.127E+04 1.146E+04 -190.
1.00
 8 25 -7.00 1.000E-07 1.038E+04 1.057E+04 -183.
1.00
 9 19 -6.52 3.000E-07 9.828E+03 9.347E+03 481.
1.00
10 11 -6.00 1.000E-06 7.918E+03 7.875E+03 42.6
1.00
11 6 -5.52 3.000E-06 6.351E+03 6.801E+03 -450.
1.00
12 3 -5.00 1.000E-05 6.135E+03 6.081E+03 53.9
1.00
13 2 -4.00 1.000E-04 5.724E+03 5.569E+03 155.
1.00

```

```

---- CURVE # 2: DOPAMI
DATA SPLINE Log X ORIG. X Y Y CALC
RESIDUAL WEIGHT
# KNOT #
-----
-----
14 38 Neg. Inf. .000 1.256E+04 1.261E+04 -53.5
1.00
15 36 -9.00 1.000E-09 1.242E+04 1.226E+04 156.
1.00
16 34 -8.00 1.000E-08 1.174E+04 1.184E+04 -95.9
1.00
17 30 -7.52 3.000E-08 1.129E+04 1.150E+04 -212.
1.00
18 29 -7.00 1.000E-07 1.133E+04 1.099E+04 345.

```

1.00							
19	26	-6.52	3.000E-07	1.033E+04	1.038E+04	-48.8	
1.00							
20	22	-6.00	1.000E-06	9.443E+03	9.553E+03	-110.	
1.00							
21	18	-5.52	3.000E-06	8.610E+03	8.695E+03	-84.5	
1.00							
22	14	-5.00	1.000E-05	7.853E+03	7.710E+03	143.	
1.00							
23	8	-4.00	1.000E-04	5.984E+03	6.023E+03	-38.7	
1.00							

---- CURVE # 3: EPINEP

DATA SPLINE	Log X	ORIG. X	Y	Y CALC	RESIDUAL	WEIGHT
#	KNOT #					

24	38	Neg. Inf.	.000	1.294E+04	1.289E+04	52.3
1.00						
25	32	-8.00	1.000E-08	1.209E+04	1.213E+04	-36.3
1.00						
26	28	-7.00	1.000E-07	1.125E+04	1.132E+04	-75.3
1.00						
27	24	-6.52	3.000E-07	1.063E+04	1.076E+04	-133.
1.00						
28	21	-6.00	1.000E-06	1.024E+04	1.002E+04	221.
1.00						
29	17	-5.52	3.000E-06	9.282E+03	9.259E+03	22.8
1.00						
30	13	-5.00	1.000E-05	8.526E+03	8.405E+03	121.
1.00						
31	10	-4.52	3.000E-05	7.560E+03	7.669E+03	-109.
1.00						
32	7	-4.00	1.000E-04	6.788E+03	6.970E+03	-182.
1.00						
33	4	-3.00	1.000E-03	6.140E+03	6.021E+03	119.
1.00						

---- CURVE # 4: NOREPI

DATA SPLINE	Log X	ORIG. X	Y	Y CALC
-------------	-------	---------	---	--------

RESIDUAL WEIGHT

KNOT

```

-----
-----
 34 38 Neg. Inf. .000 1.292E+04 1.298E+04 -54.9
1.00
 35 33 -8.00 1.000E-08 1.242E+04 1.229E+04 129.
1.00
 36 29 -7.00 1.000E-07 1.163E+04 1.157E+04 52.2
1.00
 37 27 -6.52 3.000E-07 1.096E+04 1.106E+04 -110.
1.00
 38 23 -6.00 1.000E-06 1.020E+04 1.038E+04 -186.
1.00
 39 20 -5.52 3.000E-06 9.799E+03 9.670E+03 129.
1.00
 40 15 -5.00 1.000E-05 8.905E+03 8.848E+03 57.0
1.00
 41 12 -4.52 3.000E-05 8.202E+03 8.119E+03 82.6
1.00
 42 9 -4.00 1.000E-04 7.289E+03 7.405E+03 -116.
1.00
 43 5 -3.00 1.000E-03 6.416E+03 6.398E+03 17.8
1.00

```

```

=====
==
=====

```

Finally, the common spline table is printed, together with the fitted Y at each data point, the first and second derivatives of the spline and the effective weight applied to that data point. Note that there are fewer than 43 points, as some were merged. Also, the weights now depend on the relative degree of vertical scaling (A-D) applied to each curve. In the original coordinates, all weights were 1.00. The vertical scaling has changed these values. By design, the points corresponding to the template curve receive very high weights. Since the logistic transform was used, all X values fall between 0 and 1. This will not be true in general. Use of the automatic template has caused all the Y values to lie between 0 and 1. Again, this is not true in general.

1/25/1989 10:40: 9

---- SPLINE FUNCTIONS ----

=====

==

=====

KNOT NORMALIZED NORMALIZED NORMALIZED NORMALIZED Y'
Y" WEIGHT
X Y Y CALC RESIDUAL

1	1.000E-06	.000	-4.669E-07	4.669E-07	1.00	
.000	5.207E+11					
2	1.902E-02	4.159E-02	1.902E-02	2.257E-02	1.00	
3.043E-06	4.746E+07					
3	9.342E-02	.101	9.342E-02	7.830E-03	1.00	
6.742E-05	4.746E+07					
4	.122	.137	.122	1.517E-02	1.00	
9.860E-05	6.109E+07					
5	.141	.143	.141	2.322E-03	1.00	
1.319E-04	5.865E+07					
6	.198	.133	.198	-6.536E-02	1.00	
2.352E-04	4.746E+07					
7	.243	.220	.243	-2.323E-02	1.00	
2.248E-04	6.109E+07					
8	.255	.250	.255	-4.378E-03	1.00	
2.110E-04	7.809E+07					
9	.272	.257	.272	-1.518E-02	1.00	
1.863E-04	5.865E+07					
10	.332	.318	.332	-1.399E-02	1.00	
6.766E-05	6.109E+07					
11	.354	.360	.354	6.183E-03	1.00	
1.298E-05	4.746E+07					
12	.366	.376	.366	1.079E-02	1.00	-
1.456E-05	5.865E+07					
13	.426	.442	.426	1.551E-02	1.00	-
1.320E-04	6.109E+07					
14	.446	.462	.446	1.615E-02	1.00	-
1.570E-04	7.809E+07					

15	.461	.468	.461	7.436E-03	1.00	-
	1.642E-04	5.865E+07				
16	.500	.500	.500	-7.749E-07	1.00	-
	1.717E-04	5.207E+11				
17	.536	.539	.536	2.912E-03	1.00	-
	1.879E-04	6.109E+07				
18	.557	.547	.557	-9.566E-03	1.00	-
	1.950E-04	7.809E+07				
19	.567	.637	.567	6.986E-02	1.00	-
	2.036E-04	4.746E+07				
20	.568	.585	.568	1.679E-02	1.00	-
	2.027E-04	5.865E+07				
21	.633	.661	.633	2.826E-02	1.00	-
	7.303E-05	6.109E+07				
22	.654	.642	.654	-1.243E-02	1.00	-
	6.344E-06	7.809E+07				
23	.661	.637	.661	-2.430E-02	1.00	
	1.108E-05	5.865E+07				
24	.728	.711	.728	-1.705E-02	1.00	
	1.156E-04	6.109E+07				
25	.744	.718	.744	-2.659E-02	1.00	
	1.300E-04	4.746E+07				
26	.747	.742	.747	-5.520E-03	1.00	
	1.301E-04	7.809E+07				
27	.750	.736	.750	-1.432E-02	1.00	
	1.294E-04	5.865E+07				
28	.800	.790	.800	-9.638E-03	1.00	
	9.019E-05	6.109E+07				
29	.817	.842	.817	2.520E-02	1.00	
	7.054E-05	1.367E+08				
30	.874	.850	.874	-2.402E-02	1.00	
	1.336E-04	7.809E+07				
31	.875	.847	.875	-2.751E-02	1.00	
	1.336E-04	4.746E+07				
32	.903	.898	.903	-4.639E-03	1.00	
	1.055E-04	6.109E+07				
33	.910	.927	.910	1.687E-02	1.00	
	9.620E-05	5.865E+07				
34	.913	.902	.913	-1.085E-02	1.00	
	9.492E-05	7.809E+07				
35	.939	.922	.939	-1.733E-02	1.00	
	6.588E-05	4.746E+07				
36	.961	.979	.961	1.769E-02	1.00	

3.037E-05	7.809E+07				
37	.988	.983	.988	-5.202E-03	1.00
1.070E-05	4.746E+07				
38	1.00	1.00	1.00	2.587E-06	1.00
6.072E-22	5.210E+11				

We include a graph of the data and the fitted curves. Each curve in the plot has the same "shape". The curves differ in their horizontal location and stretch. Plotting the data is accomplished by selecting Plot from the Flexifit Menu and then selecting the desired plot from the sub menu. The Plot Set Up Dialog will be displayed. You simply hit return or click on the Plot button to create the plot with all the default values.

The Common Shape and Common Shape Linearized plots are also created using menu selections.

5 IBM-PC Operation

Program Options ("Switches")

Several options are provided as "switches" which are appended to the DOS command which invokes the program. As many "switches" may be chosen and strung together as desired. The available switches are:

Switches which control the degree of smoothing:

- /C Percentage smoothing is not to be specified. Without this switch, you will be asked to specify the percentage smoothing.
- /I Inflection points -- number of these is used to determining smoothing
- /M Require curves to be monotonic (also controls degree of smoothing)

- /P Value of l^* (lambda-star) to be specified. This is the quickest mode, if you know the appropriate value for l^* . Try the value 107 for maximal smoothing, 1 for somewhere around 50% smoothing. Generally, use /X rather than /P if you wish to use small values of l^*
- /F Starting value for l^* to be specified
- /X Final value of l^* to be specified
- /A Use all available criteria for controlling degree of smoothing
- /B Bell-shaped curves -- use only if you expect non-monotonic "bell"-like shape

Utility switches:

- /D Debug -- for program development only
- /U No session file to be created (Session file created by default)
- /T An external template curve is to be provided by user
- /V Print the full covariance table
- /W Specify values of A_0 , A_1 , and A_2 for weighted fit. The model for the variance is $\text{Var}(Y) = A_0 + A_1 * Y^{A_2}$, where Y is the fitted value of the response in the original coordinate system. Weights are set inversely proportional to the modeled variance value.

For example, the DOS command

```
C>FLEXIFIT/X/M/V
```

invokes the program FLEXIFIT with the switches X, M and V (final l^* , monotonicity, and print covariance table).

Logistic Curves Example

We now describe a session of FLEXIFIT using the data file SAMPLE from a previous section. The data represent inhibition of binding of a radioactively labeled drug by four unlabeled drugs. This identical data set is described in the ALLFIT User's Guide (November 1988). This run of FLEXIFIT is arranged so that identical results are obtained. Thus, this first approach is a good starting place for previous ALLFIT users. Not described here, but in common with ALLFIT, FLEXIFIT has the potential to

constrain the parameters to constant values or to values taken by other parameters. Thus, since the amount of inhibition in the presence of a zero dose of drug is theoretically identical, it might be appropriate to constrain $A1 = A2 = A3 = A4$. One can also attempt the constraint $D1 = D2 = D3 = D4$ to test if the maximal inhibitions for all four drugs are the same.

We run FLEXIFIT without any switches. By default, the program will prompt for Percentage Smoothing. For the first run we specify 100%. The program will also ask if a logistic transform is to be applied. We shall use it here. We also choose to take logs of the data. This combination is identical to applying the four-parameter logistic function to the data, as in ALLFIT. In the session, underlining indicates what the user has typed, all the rest is what the computer has output.

```
C>FLEXIFIT
```

```
----- FLEXIFIT v. 2.1 ----- 2/16/1989 10:24:19
```

```
ENTER ONE DATA FILE NAME PER LINE, END WITH EMPTY LINE.
```

```
DATA FILE NAME -->SAMPLE.ALL
```

```
DATA FILE NAME --> <CR>
```

```
DO WE NEED TO TAKE LOGS OF DATA (Y/N) [Y] ? <CR>
```

```
DO WE USE LOGISTIC TRANSFORM (Y/N) [Y] ? <CR>
```

```
ENTER MINIMUM % SMOOTHING REQUIRED [ 0 ] -->100
```

```
TEMPLATE CURVE:
```

```
    CURVE: BLTIN3 --- CURVE # 0 3-POINT BUILT-IN TEMPLATE
```

```
***** DATA FILE(S) OPENED *****
```

```
DATA FILE: SAMPLE.ALL
```

```
    CURVE: APOMOR --- CURVE # 1 HAS 10 POINTS.
```

```
    CURVE: DOPAMI --- CURVE # 2 HAS 10 POINTS.
```

```
    CURVE: EPINEP --- CURVE # 3 HAS 10 POINTS.
```

```
    CURVE: NOREPI --- CURVE # 4 HAS 10 POINTS.
```

```
***** VALUES OF CONSTRAINTS *****
```

```
TRANSFORMATION FORMULA          : LOGISTIC
```

```
TAKE LOG OF X-DATA              : YES
```

```
TEMPLATE CURVE INCLUDED         : BLTIN3
```

MONOTONICITY NOT REQUIRED
 MAXIMUM # OF INFLECTION POINTS : 44
 LAMBDA* ADJUSTED BY PROGRAM
 PERCENTAGE SMOOTHING SPECIFIED : 100.0
 COEFFICIENTS OF VARIANCE FUNCTION :
 A0 : 1.00000
 A1 : .000000
 A2 : .000000

We have specified only one input filename, although more are possible. The file SAMPLE.ALL contains four data curves, each with 10 points. The computer will take logs of the x-data by default, use the logistic transform, and fit a 100% smooth spline, i.e. a straight line. A 3 point template curve (BLTIN#, Curve #0) is automatically generated and the corresponding values for parameters A0, B0, C0 and D0 will be fixed at pre-determined values. We allow for up to 44 inflection points, the maximum possible with 44 points. Finally, by default we are using an unweighted fit. If the switch /W had been specified in the command line, the user could input values for A0, A1, and A2 which determine the variance model. We now continue with the session.

SHARED PARAMETERS : <CR>
 NONE
 CONSTANT PARAMETERS : <CR>
 NONE

CURVE		Ai	Bi	Ci
Di				
TITLE	#	-----		

BLTIN3	0	7.21600E+05 C	1.0000	C .00000 C
.00000	C			
APOMOR	1	12570.	-.40000	-6.5000
5724.0				
DOPAMI	2	12556.	-.40000	-6.5000
5984.0				
EPINEP	3	12940.	-.40000	-5.5000
6140.0				

NOREPI 4 12923. -.40000 -5.5000
6416.0

REMARK: C:CONSTANT; S:SHARED.

ARE THESE PARAMETERS O.K.? (Y/N) --> Y

In this session we do not make use of the Shared Parameters, or Constant Parameters features. These features work exactly as they do in ALLFIT, and allow a variety of constraints to be added to the parameters. The parameters for Curve #0, the template, are automatically set to constraint values. Finally, you are given the option to reset the computer-derived initial estimates of the parameters. In most cases, automatic estimates will work, but for more difficult problems, careful, manual readjustment of these values is required.

FIT # 1-----

ITR #	LAMBDA*	SSQ	SSQ_LR	% SMOOTHING
LSQ ITR#				
1	1.000000E+07	1.008755E+06	1.008773E+06	100.00
4				
2	1000000.	1.008602E+06	1.008773E+06	99.98
1				
3	3.162278E+06	1.008718E+06	1.008773E+06	99.99
1				
4	5.623413E+06	1.008742E+06	1.008773E+06	100.00
1				
5	4.216965E+06	1.008732E+06	1.008773E+06	100.00
1				
6	3.651741E+06	1.008725E+06	1.008773E+06	100.00
1				
7	3.398208E+06	1.008722E+06	1.008773E+06	99.99
1				
8	3.522695E+06	1.008724E+06	1.008773E+06	100.00
1				

---- FINAL LAMBDA, SSQ, AND % SMOOTHING ----

LAMBDA* 3.522695E+06
LAMBDA 3.038175E+09
FINAL WEIGHTED SSQ 1.111305E+06

FINAL SPLINE SSQ 1.008724E+06
 % SMOOTHING 100.00

1/25/1989

10:40: 9

 ***** FINAL RESULTS *****

CONDITIONS : MONOTONIC
 1 INFLECTION POINTS
 WEIGHTED SSQ IN ORIGINAL COORDINATES: 1111305.0
 WEIGHTED SSQ IN SPLINE COORDINATES : 1008724.0
 SSQ DUE TO MERGING : 102581.30
 MEAN SQUARE (MS) : 41159.430
 ROOT MEAN SQUARE (RMS) : 202.87790
 NORMALIZED CURVATURE : 2.16680900E-13

---- RESULT OF THE CURRENT FIT ----

=====
 ==

=====

Curve No.	of Weighted Residuals	Points	Sum of Sqr.	MSSD	Residuals	No. of Runs	
#					+	-	
0	3	.4693	N/A	2	1	3	+++
1	10	6.0747E+05	GOOD	5	5	5	+-----+
++	2	2.4385E+05	GOOD	3	7	7	-++++---
+-	3	1.5046E+05	GOOD	5	5	5	+-----+
-+	4	1.0952E+05	GOOD	6	4	6	-++++---
+	Total	43	1.1113E+06	GOOD	21	22	18 +-----+

-----++-

(See Macintosh chapter for an explanation of the output.)

1/25/1989 10:40: 9

---- PARAMETER VALUES ----

=====
==

=====
Curve Parameter Value Standard RELATIVE
Shared Error Error
Name Error Error
Groups

BLTIN3	A0	7.2160E+05		
APOMOR	A1	1.2327E+04	150.9	1.2%
DOPAMI	A2	1.2610E+04	189.4	1.5%
EPINEP	A3	1.2888E+04	203.2	1.5%
NOREPI	A4	1.2978E+04	202.4	1.5%

BLTIN3	B0	1.000		
APOMOR	B1	-.7255	7.6341E-02	10.5%
DOPAMI	B2	-.3716	5.7438E-02	15.4%
EPINEP	B3	-.3652	4.5034E-02	12.3%
NOREPI	B4	-.3584	4.7306E-02	13.1%

			<----- ALLFIT 'C' ----->	<--- FLEXIFIT	
			Value		
Std. Error					
BLTIN3	C0	1.000		.0000	
APOMOR	C1	4.3614E-07	7.3051E-08	16.7%	-6.360
		7.2742E-02			
DOPAMI	C2	5.5544E-06	4.3884E-06	79.0%	-5.255

.3431
 EPINEP C3 4.4418E-06 2.0254E-06 45.6% -5.352
 .1980
 NOREPI C4 6.4479E-06 3.3727E-06 52.3% -5.191
 .2272

POTENCY RATIO

C2/C1 12.74 10.29 80.8%
 C1/C2 7.8521E-02 6.3416E-02
 C3/C1 10.18 4.947 48.6%
 C1/C3 9.8189E-02 4.7698E-02
 C4/C1 14.78 8.120 54.9%
 C1/C4 6.7640E-02 3.7150E-02

BLTIN3 DO .0000
 APOMOR D1 5437. 234.0 4.3%
 DOPAMI D2 3773. 1191. 31.5%
 EPINEP D3 5072. 535.2 10.5%
 NOREPI D4 5319. 606.1 11.3%

=====
 ==
 =====

1/25/1989 10:40: 9

---- ORIGINAL DATA ----

=====
 ==
 =====

---- CURVE #0 TEMPLATE: BLTIN3
 DATA SPLINE Log X ORIG. X Y Y CALC
 RESIDUAL WEIGHT
 # KNOT #

```

-----
  1  1  -6.00   1.000E-06  .000   -.337   .337
1.00
  2 16   .000    1.00   3.608E+05  3.608E+05  -.563
1.00
  3 38   6.00   1.000E+06  7.216E+05  7.216E+05  .250
1.00

```

```

---- CURVE # 1: APOMOR
DATA SPLINE Log X   ORIG. X   Y       Y CALC
RESIDUAL WEIGHT
# KNOT #

```

```

-----
  4 38 Neg. Inf. .000   1.257E+04  1.233E+04  243.
1.00
  5 37  -9.00   1.000E-09  1.221E+04  1.224E+04  -35.8
1.00
  6 35  -8.00   1.000E-08  1.179E+04  1.191E+04  -119.
1.00
  7 31  -7.52   3.000E-08  1.127E+04  1.146E+04  -190.
1.00
  8 25  -7.00   1.000E-07  1.038E+04  1.057E+04  -183.
1.00
  9 19  -6.52   3.000E-07  9.828E+03  9.347E+03  481.
1.00
 10 11  -6.00   1.000E-06  7.918E+03  7.875E+03  42.6
1.00
 11  6  -5.52   3.000E-06  6.351E+03  6.801E+03  -450.
1.00
 12  3  -5.00   1.000E-05  6.135E+03  6.081E+03  53.9
1.00
 13  2  -4.00   1.000E-04  5.724E+03  5.569E+03  155.
1.00

```

```

---- CURVE # 2: DOPAMI
DATA SPLINE Log X   ORIG. X   Y       Y CALC
RESIDUAL WEIGHT
# KNOT #

```

```

-----
 14 38 Neg. Inf. .000   1.256E+04  1.261E+04  -53.5
1.00

```

15	36	-9.00	1.000E-09	1.242E+04	1.226E+04	156.
1.00						
16	34	-8.00	1.000E-08	1.174E+04	1.184E+04	-95.9
1.00						
17	30	-7.52	3.000E-08	1.129E+04	1.150E+04	-212.
1.00						
18	29	-7.00	1.000E-07	1.133E+04	1.099E+04	345.
1.00						
19	26	-6.52	3.000E-07	1.033E+04	1.038E+04	-48.8
1.00						
20	22	-6.00	1.000E-06	9.443E+03	9.553E+03	-110.
1.00						
21	18	-5.52	3.000E-06	8.610E+03	8.695E+03	-84.5
1.00						
22	14	-5.00	1.000E-05	7.853E+03	7.710E+03	143.
1.00						
23	8	-4.00	1.000E-04	5.984E+03	6.023E+03	-38.7
1.00						

---- CURVE # 3: EPINEP

DATA SPLINE	Log X	ORIG. X	Y	Y CALC	RESIDUAL	WEIGHT
#	KNOT #					

24	38	Neg. Inf.	.000	1.294E+04	1.289E+04	52.3
1.00						
25	32	-8.00	1.000E-08	1.209E+04	1.213E+04	-36.3
1.00						
26	28	-7.00	1.000E-07	1.125E+04	1.132E+04	-75.3
1.00						
27	24	-6.52	3.000E-07	1.063E+04	1.076E+04	-133.
1.00						
28	21	-6.00	1.000E-06	1.024E+04	1.002E+04	221.
1.00						
29	17	-5.52	3.000E-06	9.282E+03	9.259E+03	22.8
1.00						
30	13	-5.00	1.000E-05	8.526E+03	8.405E+03	121.
1.00						
31	10	-4.52	3.000E-05	7.560E+03	7.669E+03	-109.
1.00						
32	7	-4.00	1.000E-04	6.788E+03	6.970E+03	-182.
1.00						

33 4 -3.00 1.000E-03 6.140E+03 6.021E+03 119.
1.00

---- CURVE # 4: NOREPI

DATA SPLINE Log X ORIG. X Y Y CALC
RESIDUAL WEIGHT
KNOT

34 38 Neg. Inf. .000 1.292E+04 1.298E+04 -54.9
1.00
35 33 -8.00 1.000E-08 1.242E+04 1.229E+04 129.
1.00
36 29 -7.00 1.000E-07 1.163E+04 1.157E+04 52.2
1.00
37 27 -6.52 3.000E-07 1.096E+04 1.106E+04 -110.
1.00
38 23 -6.00 1.000E-06 1.020E+04 1.038E+04 -186.
1.00
39 20 -5.52 3.000E-06 9.799E+03 9.670E+03 129.
1.00
40 15 -5.00 1.000E-05 8.905E+03 8.848E+03 57.0
1.00
41 12 -4.52 3.000E-05 8.202E+03 8.119E+03 82.6
1.00
42 9 -4.00 1.000E-04 7.289E+03 7.405E+03 -116.
1.00
43 5 -3.00 1.000E-03 6.416E+03 6.398E+03 17.8
1.00

=====
==
=====

We include a graph of the data and the fitted curves. Each curve in the plot has the same "shape". The curves differ in their horizontal location and stretch. Plotting the data is accomplished with program GRAFIT, described in a later section.

Here is the graph of the linearized common spline.

1/25/1989 10:40: 9

---- SPLINE FUNCTIONS ----

=====
==

=====
KNOT NORMALIZED NORMALIZED NORMALIZED NORMALIZED Y'
Y" WEIGHT
X Y Y CALC RESIDUAL

1	1.000E-06	.000	-4.669E-07	4.669E-07	1.00
.000	5.207E+11				
2	1.902E-02	4.159E-02	1.902E-02	2.257E-02	1.00
3.043E-06	4.746E+07				
3	9.342E-02	.101	9.342E-02	7.830E-03	1.00
6.742E-05	4.746E+07				
4	.122	.137	.122	1.517E-02	1.00
9.860E-05	6.109E+07				
5	.141	.143	.141	2.322E-03	1.00
1.319E-04	5.865E+07				
6	.198	.133	.198	-6.536E-02	1.00
2.352E-04	4.746E+07				
7	.243	.220	.243	-2.323E-02	1.00
2.248E-04	6.109E+07				
8	.255	.250	.255	-4.378E-03	1.00
2.110E-04	7.809E+07				
9	.272	.257	.272	-1.518E-02	1.00
1.863E-04	5.865E+07				
10	.332	.318	.332	-1.399E-02	1.00
6.766E-05	6.109E+07				
11	.354	.360	.354	6.183E-03	1.00
1.298E-05	4.746E+07				

12	.366	.376	.366	1.079E-02	1.00	-
1.456E-05	5.865E+07					
13	.426	.442	.426	1.551E-02	1.00	-
1.320E-04	6.109E+07					
14	.446	.462	.446	1.615E-02	1.00	-
1.570E-04	7.809E+07					
15	.461	.468	.461	7.436E-03	1.00	-
1.642E-04	5.865E+07					
16	.500	.500	.500	-7.749E-07	1.00	-
1.717E-04	5.207E+11					
17	.536	.539	.536	2.912E-03	1.00	-
1.879E-04	6.109E+07					
18	.557	.547	.557	-9.566E-03	1.00	-
1.950E-04	7.809E+07					
19	.567	.637	.567	6.986E-02	1.00	-
2.036E-04	4.746E+07					
20	.568	.585	.568	1.679E-02	1.00	-
2.027E-04	5.865E+07					
21	.633	.661	.633	2.826E-02	1.00	-
7.303E-05	6.109E+07					
22	.654	.642	.654	-1.243E-02	1.00	-
6.344E-06	7.809E+07					
23	.661	.637	.661	-2.430E-02	1.00	
1.108E-05	5.865E+07					
24	.728	.711	.728	-1.705E-02	1.00	
1.156E-04	6.109E+07					
25	.744	.718	.744	-2.659E-02	1.00	
1.300E-04	4.746E+07					
26	.747	.742	.747	-5.520E-03	1.00	
1.301E-04	7.809E+07					
27	.750	.736	.750	-1.432E-02	1.00	
1.294E-04	5.865E+07					
28	.800	.790	.800	-9.638E-03	1.00	
9.019E-05	6.109E+07					
29	.817	.842	.817	2.520E-02	1.00	
7.054E-05	1.367E+08					
30	.874	.850	.874	-2.402E-02	1.00	
1.336E-04	7.809E+07					
31	.875	.847	.875	-2.751E-02	1.00	
1.336E-04	4.746E+07					
32	.903	.898	.903	-4.639E-03	1.00	
1.055E-04	6.109E+07					
33	.910	.927	.910	1.687E-02	1.00	

9.620E-05	5.865E+07				
34	.913	.902	.913	-1.085E-02	1.00
9.492E-05	7.809E+07				
35	.939	.922	.939	-1.733E-02	1.00
6.588E-05	4.746E+07				
36	.961	.979	.961	1.769E-02	1.00
3.037E-05	7.809E+07				
37	.988	.983	.988	-5.202E-03	1.00
1.070E-05	4.746E+07				
38	1.00	1.00	1.00	2.587E-06	1.00
6.072E-22	5.210E+11				

=====

==

=====

DO YOU WANT A NEW FIT (SAME CONSTRAINTS) (Y/N) [N] ?<CR>
 THIS SESSION WAS RECORDED IN FLEX1 .SES
 DO YOU WANT TO FIT OTHER DATA (Y/N) [N] ?<CR>

1/25/1989 10:42:31

Generalized Logistic Example

We again run FLEXIFIT without any switches. To explore the potential for a better fit using splines, we specify 50% smoothing. This will require that the residual sum-of-squares be reduced 50% from what would be obtained using a straight line. Whether or not this results in a significant improvement in fit can be evaluated from the graphs, the runs test, or via the use of a test based on the apparent degrees of freedom. This latter calculation will be implemented in forthcoming versions of FLEXIFIT.

N.B. The RMS error and Standard Errors for parameters in this example have not been corrected for the effect of 50% smoothing. (This will be included in future versions of FLEXIFIT). Thus, these values should be seen as under-estimates of the true RMS and Standard Errors.

----- FLEXIFIT v. 2.1 ----- 1/25/1989 10:42:36

TEMPLATE CURVE:

CURVE: BLTIN3 --- CURVE # 0 3-POINT BUILT-IN TEMPLATE

***** DATA FILE(S) OPENED *****

DATA FILE: SAMPLE.ALL

CURVE: APOMOR --- CURVE # 1 HAS 10 POINTS.

CURVE: DOPAMI --- CURVE # 2 HAS 10 POINTS.

CURVE: EPINEP --- CURVE # 3 HAS 10 POINTS.

CURVE: NOREPI --- CURVE # 4 HAS 10 POINTS.

***** VALUES OF CONSTRAINTS *****

TRANSFORMATION FORMULA : LOGISTIC

TAKE LOG OF X-DATA : YES

TEMPLATE CURVE INCLUDED : BLTIN3

MONOTONICITY NOT REQUIRED

MAXIMUM # OF INFLECTION POINTS : 44

LAMBDA* ADJUSTED BY PROGRAM

PERCENTAGE SMOOTHING SPECIFIED : 50.0

COEFFICIENTS OF VARIANCE FUNCTION :

A0 : 1.00000

A1 : .000000

A2 : .000000

SHARED PARAMETERS :

NONE

CONSTANT PARAMETERS :

NONE

CURVE		Ai	Bi	Ci
Di				
TITLE	#	-----		

BLTIN3	0	7.21600E+05 C	1.0000	C .00000 C
.00000	C			
APOMOR	1	12570.	-.40000	-6.5000
5724.0				
DOPAMI	2	12556.	-.40000	-6.5000
5984.0				
EPINEP	3	12940.	-.40000	-5.5000
6140.0				
NOREPI	4	12923.	-.40000	-5.5000
6416.0				

REMARK: C:CONSTANT; S:SHARED.

ARE THESE PARAMETERS O.K.? (Y/N) --> Y

FIT # 1-----

ITR #	LAMBDA*	SSQ	SSQ_LR	% SMOOTHING
LSQ ITR#				
1	1.000000E+07	1.008755E+06	1.008773E+06	100.00
4				
2	1000000.	1.008602E+06	1.008773E+06	99.98
1				
3	100000.	1.007096E+06	1.008773E+06	99.83
1				
4	10000.0	988773.	1.005812E+06	98.31
3				
5	1000.00	924484.	1.102685E+06	83.84
9				
6	100.000	702666.	1.699977E+06	41.33
22				
7	316.228	811544.	1.314346E+06	61.75
11				
8	177.828	754182.	1.500709E+06	50.26
16				
9	133.352	727756.	1.601820E+06	45.43
14				
10	153.993	736894.	1.548132E+06	47.60
10				
11	165.482	747425.	1.526149E+06	48.97
7				
12	171.544	750823.	1.513362E+06	49.61
5				

---- FINAL LAMBDA, SSQ, AND % SMOOTHING ----

LAMBDA* 177.828
LAMBDA 139456.
FINAL WEIGHTED SSQ 798971.
FINAL SPLINE SSQ 755120.
% SMOOTHING 50.63

In this run, 12 iterations were required. Some of these major iterations included as many as 22 least-squares iterations. Thus, the computing time required goes up as the percentage smoothing drops. In some instances, a notation of "not converged" will be

found for a particular iteration. In such cases, l^* will be increased, and convergence should then be obtained. The minimal degree of smoothing will then be determined by the ability of the algorithm to converge.

1/25/1989 10:42:36

 ***** FINAL RESULTS *****

CONDITIONS : MONOTONIC
 1 INFLECTION POINTS
 WEIGHTED SSQ IN ORIGINAL COORDINATES: 798971.10
 WEIGHTED SSQ IN SPLINE COORDINATES : 755119.60
 SSQ DUE TO MERGING : 43851.560
 MEAN SQUARE (MS) : 29591.520
 ROOT MEAN SQUARE (RMS) : 172.02190
 NORMALIZED CURVATURE : 1.86871100E-05

---- RESULT OF THE CURRENT FIT ----

=====
 ==

=====							
Curve No.	No. of Weighted Residuals	Sum of Sqr.	MSSD	Residuals	No. of Runs		
#	Points			+ -			

0	3	.1293	N/A	1 2	2	--+	
1	10	3.8182E+05	GOOD	5 5	5	+-----+-	
++							
2	10	2.6175E+05	GOOD	4 6	6	-+---+---	
++							
3	10	6.8763E+04	GOOD	4 6	7	+-----+-	
-+							
4	10	8.6645E+04	GOOD	5 5	8	-+-----+-	
+							
Total	43	7.9897E+05	GOOD	19 24	19	-+-----+-	
-+-----+-							

---+---

+---+---+

=====
==
=====

1/25/1989 10:42:36

---- PARAMETER VALUES ----

=====
==
=====

Curve Shared Name Groups	Parameter	Value	Standard Error	RELATIVE Error
--------------------------------	-----------	-------	-------------------	-------------------

BLTIN3	A0	7.2160E+05		
APOMOR	A1	1.2373E+04	132.9	1.0%
DOPAMI	A2	1.2593E+04	155.6	1.2%
EPINEP	A3	1.2921E+04	174.4	1.3%
NOREPI	A4	1.2982E+04	172.9	1.3%

BLTIN3	B0	1.000		
APOMOR	B1	-.7220	6.1292E-02	8.4%
DOPAMI	B2	-.4123	4.2126E-02	10.2%
EPINEP	B3	-.3843	3.2615E-02	8.4%
NOREPI	B4	-.3785	3.3493E-02	8.8%

<----- ALLFIT 'C' -----> <---- FLEXIFIT
'C' ---->

Value

Std. Error

BLTIN3	C0	1.000			.0000
APOMOR	C1	4.4751E-07	5.9955E-08	13.4%	-6.349
		5.8184E-02			
DOPAMI	C2	3.5839E-06	1.5197E-06	42.4%	-5.443
		.1841			
EPINEP	C3	3.5923E-06	1.0137E-06	28.2%	-5.443
		.1226			
NOREPI	C4	5.5261E-06	1.7357E-06	31.4%	-5.257
		.1364			

POTENCY RATIO

C2/C1	8.051	3.580	44.5%
C1/C2	.1242	5.5236E-02	
C3/C1	8.049	2.514	31.2%
C1/C3	.1242	3.8808E-02	
C4/C1	12.37	4.223	34.1%
C1/C4	8.0852E-02	2.7608E-02	

BLTIN3	D0	.0000		
APOMOR	D1	5529.	175.7	3.1%
DOPAMI	D2	4620.	596.7	12.9%
EPINEP	D3	5485.	301.3	5.4%
NOREPI	D4	5698.	337.8	5.9%

=====
==
=====

1/25/1989 10:42:36

---- ORIGINAL DATA ----

=====
==
=====

---- CURVE #0 TEMPLATE: BLTIN3

DATA SPLINE Log X ORIG. X Y Y CALC
 RESIDUAL WEIGHT
 # KNOT #

#	KNOT #	Log X	ORIG. X	Y	Y CALC	RESIDUAL	WEIGHT
1	1	-6.00	1.000E-06	.000	6.135E-02	-6.135E-02	1.00
2	16	.000	1.00	3.608E+05	3.608E+05	-.313	1.00
3	39	6.00	1.000E+06	7.216E+05	7.216E+05	.188	1.00

---- CURVE # 1: APOMOR

DATA SPLINE Log X ORIG. X Y Y CALC
 RESIDUAL WEIGHT
 # KNOT #

#	KNOT #	Log X	ORIG. X	Y	Y CALC	RESIDUAL	WEIGHT
4	39	Neg. Inf.	.000	1.257E+04	1.237E+04	197.	1.00
5	38	-9.00	1.000E-09	1.221E+04	1.228E+04	-69.7	1.00
6	36	-8.00	1.000E-08	1.179E+04	1.190E+04	-107.	1.00
7	31	-7.52	3.000E-08	1.127E+04	1.141E+04	-138.	1.00
8	26	-7.00	1.000E-07	1.038E+04	1.051E+04	-130.	1.00
9	20	-6.52	3.000E-07	9.828E+03	9.429E+03	399.	1.00
10	12	-6.00	1.000E-06	7.918E+03	7.903E+03	15.3	1.00
11	6	-5.52	3.000E-06	6.351E+03	6.693E+03	-342.	1.00
12	3	-5.00	1.000E-05	6.135E+03	6.052E+03	82.5	1.00
13	2	-4.00	1.000E-04	5.724E+03	5.636E+03	88.3	1.00

---- CURVE # 2: DOPAMI

DATA SPLINE Log X ORIG. X Y Y CALC

RESIDUAL WEIGHT

KNOT

```

-----
-----
14 39 Neg. Inf. .000 1.256E+04 1.259E+04 -36.8
1.00
15 37 -9.00 1.000E-09 1.242E+04 1.229E+04 134.
1.00
16 35 -8.00 1.000E-08 1.174E+04 1.185E+04 -105.
1.00
17 32 -7.52 3.000E-08 1.129E+04 1.149E+04 -204.
1.00
18 29 -7.00 1.000E-07 1.133E+04 1.096E+04 370.
1.00
19 25 -6.52 3.000E-07 1.033E+04 1.034E+04 -8.36
1.00
20 22 -6.00 1.000E-06 9.443E+03 9.568E+03 -125.
1.00
21 18 -5.52 3.000E-06 8.610E+03 8.753E+03 -143.
1.00
22 13 -5.00 1.000E-05 7.853E+03 7.724E+03 129.
1.00
23 7 -4.00 1.000E-04 5.984E+03 5.979E+03 5.16
1.00

```

---- CURVE # 3: EPINEP

DATA SPLINE Log X ORIG. X Y Y CALC
RESIDUAL WEIGHT
KNOT

```

-----
-----
24 39 Neg. Inf. .000 1.294E+04 1.292E+04 18.7
1.00
25 33 -8.00 1.000E-08 1.209E+04 1.212E+04 -28.0
1.00
26 28 -7.00 1.000E-07 1.125E+04 1.128E+04 -35.0
1.00
27 24 -6.52 3.000E-07 1.063E+04 1.072E+04 -92.6
1.00
28 21 -6.00 1.000E-06 1.024E+04 1.005E+04 192.
1.00
29 17 -5.52 3.000E-06 9.282E+03 9.332E+03 -50.1

```

1.00
 30 14 -5.00 1.000E-05 8.526E+03 8.439E+03 86.9
 1.00
 31 10 -4.52 3.000E-05 7.560E+03 7.608E+03 -48.3
 1.00
 32 8 -4.00 1.000E-04 6.788E+03 6.870E+03 -81.6
 1.00
 33 4 -3.00 1.000E-03 6.140E+03 6.098E+03 42.5
 1.00

---- CURVE # 4: NOREPI
 DATA SPLINE Log X ORIG. X Y Y CALC
 RESIDUAL WEIGHT
 # KNOT #

 34 39 Neg. Inf. .000 1.292E+04 1.298E+04 -59.5
 1.00
 35 34 -8.00 1.000E-08 1.242E+04 1.228E+04 141.
 1.00
 36 30 -7.00 1.000E-07 1.163E+04 1.154E+04 84.1
 1.00
 37 27 -6.52 3.000E-07 1.096E+04 1.103E+04 -73.9
 1.00
 38 23 -6.00 1.000E-06 1.020E+04 1.039E+04 -198.
 1.00
 39 19 -5.52 3.000E-06 9.799E+03 9.753E+03 46.4
 1.00
 40 15 -5.00 1.000E-05 8.905E+03 8.919E+03 -13.9
 1.00
 41 11 -4.52 3.000E-05 8.202E+03 8.108E+03 93.6
 1.00
 42 9 -4.00 1.000E-04 7.289E+03 7.309E+03 -20.2
 1.00
 43 5 -3.00 1.000E-03 6.416E+03 6.416E+03
 6.348E-03 1.00

=====
 ==
 =====

1/25/1989 10:42:36

---- SPLINE FUNCTIONS ----

=====
==
=====

KNOT NORMALIZED NORMALIZED NORMALIZED NORMALIZED Y'
Y" WEIGHT
X Y Y CALC RESIDUAL

1	1.000E-06	.000	8.502E-08	-8.502E-08	.792	
.000	5.207E+11					
2	1.973E-02	2.853E-02	1.562E-02	1.290E-02	.792	-
1.253E-02	4.685E+07					
3	9.596E-02	8.857E-02	7.652E-02	1.205E-02	.814	
.600	4.685E+07					
4	.103	8.808E-02	8.237E-02	5.715E-03	.819	
.715	5.530E+07					
5	.123	9.854E-02	9.854E-02	8.946E-07	.837	
1.12	5.306E+07					
6	.202	.120	.170	-5.004E-02	.991	
2.76	4.685E+07					
7	.202	.171	.170	6.477E-04	.991	
2.76	6.356E+07					
8	.218	.175	.186	-1.097E-02	1.03	
2.56	5.530E+07					
9	.251	.218	.221	-2.768E-03	1.11	
1.88	5.306E+07					
10	.307	.279	.286	-6.499E-03	1.17	
.574	5.530E+07					
11	.345	.344	.331	1.285E-02	1.18	-
.516	5.306E+07					
12	.359	.349	.347	2.241E-03	1.17	-
.769	4.685E+07					
13	.396	.405	.389	1.619E-02	1.13	-
1.40	6.356E+07					
14	.403	.409	.397	1.169E-02	1.12	-
1.41	5.530E+07					
15	.444	.440	.442	-1.908E-03	1.06	-
1.13	5.306E+07					

16	.500	.500	.500	-4.172E-07	1.01	-
.817	5.207E+11					
17	.517	.511	.517	-6.743E-03	.996	-
.775	5.530E+07					
18	.518	.500	.518	-1.794E-02	.995	-
.778	6.356E+07					
19	.558	.563	.557	6.371E-03	.950	-
1.53	5.306E+07					
20	.572	.628	.570	5.824E-02	.927	-
1.74	4.685E+07					
21	.620	.639	.613	2.585E-02	.871	-
.536	5.530E+07					
22	.629	.605	.621	-1.564E-02	.868	-
.166	6.356E+07					
23	.656	.617	.645	-2.720E-02	.876	-
.691	5.306E+07					
24	.722	.692	.704	-1.246E-02	.943	-
1.36	5.530E+07					
25	.735	.716	.717	-1.049E-03	.961	-
1.37	6.356E+07					
26	.747	.709	.728	-1.896E-02	.977	-
1.36	4.685E+07					
27	.751	.722	.732	-1.015E-02	.982	-
1.31	5.306E+07					
28	.798	.775	.780	-4.709E-03	1.02	-
.306	5.530E+07					
29	.814	.842	.796	4.643E-02	1.02	-
7.898E-02	6.356E+07					
30	.820	.814	.802	1.155E-02	1.02	-
3.336E-02	5.306E+07					
31	.876	.839	.859	-2.017E-02	1.06	-
1.49	4.685E+07					
32	.878	.836	.862	-2.556E-02	1.07	-
1.52	6.356E+07					
33	.906	.888	.892	-3.766E-03	1.11	-
1.22	5.530E+07					
34	.916	.923	.904	1.932E-02	1.12	-
1.08	5.306E+07					
35	.919	.893	.907	-1.316E-02	1.12	-
1.09	6.356E+07					
36	.940	.915	.930	-1.560E-02	1.14	-
.860	4.685E+07					
37	.967	.978	.962	1.687E-02	1.16	-

.279 6.356E+07
38 .988 .976 .986 -1.019E-02 1.16
.154 4.685E+07
39 1.00 1.00 1.00 1.694E-06 1.16
6.234E-18 5.209E+11

=====
==
=====

DO YOU WANT A NEW FIT (SAME CONSTRAINTS) (Y/N) [N] ?<CR>
THIS SESSION WAS RECORDED IN FLEX2 .SES
DO YOU WANT TO FIT OTHER DATA (Y/N) [N] ?<CR>

1/25/1989 10:46:10

Creating Plots on the IBM-PC

Plotting is performed on the IBM-PC using a separate program called GRAFIT. This program is distributed on the same floppy diskette with the FLEXIFIT program.

The GRAFIT program plots the data written to .grf files by FLEXIFIT.

GRAFIT has been designed to work on machines with one of the following graphics cards.

Hercules
CGA
EGA
VGA
MCGA

The type of video card is automatically detected by the program.

GRAFIT supports laser printers that use the picture drawing language PostScript. The best hard copy plots are achieved using a laser printer. Dot matrix printers are also supported. In previous versions of GRAFIT, the plots were produced at screen resolution. The software has now takes advantage of the higher resolution available on 9 pin dot matrix printers. GRAFIT uses a printing protocol which works on most dot matrix printers.

In order to support still more printers, GRAFIT is capable of producing a Lotus 123 .PIC file. This file can be printed using the print graph program distributed with Lotus 123 version 2.

The GRAFIT program exits as a single executable file. You may run it from a floppy drive or copy it to your hard disk.

Start the program by typing grafit at the DOS prompt.

```
> grafit
```

You will be prompted for a .grf file name.

```
Graphics file name (no extension) ? gsample
```

(Don't enter the .grf file extension, it is assumed.)
The file gsample.grf, found on the Flexifit distribution diskette, will be used to demonstrate GRAFIT operation.

The following plot will be displayed on the screen.

The line at the bottom of the screen contains 7 different commands...

```
C:Change N:Next F:First K:Key P:Print R:Restart Q:Quit
```

A command is entered by typing the first letter of the command. No return is necessary. These commands are only active while a plot is on the screen and the commands are displayed at the bottom of the screen.

Change Command

Type the letter 'C', the following screen will be displayed.

Graph Set Up Screen

```
Label 1: ALLFIT-PC: GSAMPLE.ALL
```

Label 2: FIT # 1 SESSION: GSAMPLE.SES

Title: title

X label: x

Y label: y

x min: -1.0000E+01

x max: -3.0000E+00

y min: 5.0000E+03

y max: 1.3000E+04

X divisions (1-15): 2

Y divisions (1-15): 2

Size of plotted points (1-10): 3

F2:Accept new values Esc:Cancel

You can change the plot labels or axis limits. The number of tick marks on the X and Y axis can be adjusted along with the size of the symbols used to plot the data points. Strike the Esc key to ignore your changes and return to the plot. F2 will register your changes and display the updated plot.

Next Command

To display the next plot in the file, type the letter 'N'. The graph file will be read and the next plot will be displayed on the screen. If no more plots are in the file a message to this effect will be displayed.

First Command

You can display the first plot in the file at any time by typing the letter 'F'.

Key (Legend) Command

After entering 'K' from the keyboard you will be presented with the following screen.

Key Set Up Screen

Write curve labels? (Y/N): N

Name of curve 1: Draw curve 1? (Y/N): Y
Name of curve 2: Draw curve 2? (Y/N): Y

F2:Accept new values Esc:Cancel

You may enter a label for each curve in the plot. Initially the key or legend is not displayed. Display the key by answering 'Y' to the question 'Write curve labels?' . The key consisting of the curve labels and associated symbols will be displayed to the right of the plot. Be careful when printing not to position the plot too far to the right as this may push the key off of the page.

You may choose to display or not display any particular curve. Answering yes to the question 'Draw curve 1?' will cause the curve to be drawn; answer no and the curve will not be drawn.

Print Command

GRAFIT supports several methods of printing your plots. Printing may be done directly to a 9 pin dot matrix printer. Alternately a PostScript file may be created and later sent to a PostScript printer. Lastly a Lotus 123 .PIC format file may be created. All of these methods are initiated from the print screen.

While a plot is displayed on the screen, entering 'P' will result in the display of the print menu.

Printer Set Up Screen

These coordinates alter the size and position of the plot box. The lower left hand corner of the page is (0,0).

Upper Left x page coordinate (inches): 2.00
Upper Left Y page coordinate (inches): 9.00
Lower Right x page coordinate (inches): 7.00
Lower Right y page coordinate (inches): 6.00

Post Script File (.ps will be appended): gsample
Lotus 123 file (.pic will be appended):..gsample

F1:PostScript F2:Save F3:Dot Matrix F4:Lotus Esc:Cancel

Creating a PostScript File

Entering F1 will result in the creation (or replacement) of the PostScript file gsample.ps. The default PostScript file name is the prefix of the .grf file. You may enter your own file name instead of using the default. To print this file you will have to exit this program (i.e. hit escape and then Quit), and then download the PostScript file gsample.ps to your laser printer. The page coordinates apply to PostScript file.

Dot Matrix Printing

Enter F3 to print the plot on a dot matrix printer attached to printer port LPT1. Once printing has started you may interrupt it by striking any key. GRAFIT supports 9 pin dot matrix printers which emulate the Epson FX series of printers. This method creates a plot at a much higher resolution than available with screen dumps. You may alter the size of the plot using the page coordinates.

Creating Lotus 123 Format Files

Entering F4 will result in the creation of the Lotus 123 .PIC file gsample.pic. (You may enter a different file name. .PIC is always appended.) This .PIC file may be printed using the Lotus 123 program Printgraph which is supplied with version 2 of Lotus 123. This enables GRAFIT to indirectly support a wide range of printers. Note, the page coordinates do not apply to this method.

F1, F2, F3 or F4 will result in your changes being saved. If you make changes in this screen and then enter Esc, without having entered F1, F2 or F3, then your changes will not be saved.

Restart Command

The restart command is used to change the .grf file.

Quit Command

Enter 'Q' to exit the program. Again, this command is only active when a plot is displayed.

Prt-Screen Keyboard Key

The prt-scr key is supported only when a plot is being displayed. Screen dumps are supported in Hercules, CGA, EGA and VGA video modes. You do not have to have the MS-DOS graphics driver loaded to use this function.

GRAFIT file format

The file format for GRAFIT is discussed below, along with a listing of a sample graph file. This file is included on the disk (gsample.grf). The number of lines in any graph file is variable, depending on the number of curves. Here is a listing of gsample.grf.

```
ALLFIT-PC: GSAMPLE.ALL FIT # 1  
SESSION: GSAMPLE.SES
```

These first two lines are graph labels. They will appear at the top of the graph.

```
2
```

This line is the number of curves.

```
10 10
```

These two numbers indicate how many points are in each curve respectively. Remember that if the number of curves were 3, then there would be three numbers here.

```
-10 -3 1
```

```
5500 13000 999.9999
```

The first line is the minimum x value, the maximum x value, and the default step value for the x axis. The second line contains the same values for the y axis.

These values are for the whole graph, not any specific curve.

-10 12570
-9 12208
-8 11789
-7.522879 11273
-7 10382
-6.522879 9828
-6 7918
-5.522879 6351
-5 6135
-4 5724

These points are the x,y pairs for the 10 points in curve 1.

-10 12556
-9 12421
-8 11743
-7.522879 11287
-7 11333
-6.522879 10328
-6 9443
-5.522879 8610
-5 7853
-4 5984

These points are the x,y pairs for the 10 points in curve 2.

55

This number is the number of lines to follow. The lines contain x,y pairs for the two endpoints of line segments. These line segments make up the curve on the graph. These are for curve 1.

-10 12299.78 -9.872727 12296.29
-9.872727 12296.29 -9.745455 12291.98
-9.745455 12291.98 -9.618182 12286.63
-9.618182 12286.63 -9.49091 12280.02
-9.49091 12280.02 -9.363637 12271.83

.
.

Not all 55 lines of numbers are shown here.

.
.

-3.509097 5503.595 -3.381825 5492.997
-3.381825 5492.997 -3.254552 5484.425
-3.254552 5484.425 -3.127279 5477.498
-3.127279 5477.498 -3.000007 5471.9

55

This number is the number of lines to follow. The lines contain x,y pairs for the two endpoints of line segments. These line segments make up the curve on the graph. These are for curve 2.

-10 12459.92 -9.872727 12443.04
-9.872727 12443.04 -9.745455 12424.3
-9.745455 12424.3 -9.618182 12403.5
-9.618182 12403.5 -9.49091 12380.42
-9.49091 12380.42 -9.363637 12354.82

.

Not all 55 lines of numbers are shown here.

.

-3.509097 5387.667 -3.381825 5248.426
-3.381825 5248.426 -3.254552 5118.917
-3.254552 5118.917 -3.127279 4998.845
3.127279 4998.845 -3.000007 4887.851

End of graph file. GRAFIT can read multiple sets of curve data. For another graph, the above format of input can be repeated in the same file (see the file session.grf on the distribution disk). Make sure not to put any blank lines in the graph file.

Index

ALLFIT C 2
ALLFIT users 20
bell shaped curve 15
Clear 10
Clipboard 11
Combine 10
common "shape" 1
constant constraint 16
Constraints 12
coprocessor chip 5
Copy 10

Cut 10
data files 7
Data Points 11
degree of smoothing 1, 29
degrees of freedom 23
families of curves 1
Fit Data 12
FLEXIFIT C 2
Flexifit Screen Output 11
format 7
goaway box 13
Graph 11
Graph Setup 11
graph window 18
graphics program 5

inflection points 15
initial estimates 21
logistic 3
minimum % smoothing 15
monotonicity 14
MSSD 23
Options 12
Parameter Values 12
Paste 10
Plot 12
Plot input 12
Print 10
Printing Plots 18

Quit Flexifit 10
Reduced View 11
Reset Parameters 12
RMS 23
Save As 10
Select All 10
session file 8
Setup 11
shared constraints 16
Standard Errors 24
template curve 15
variance model 15
Weighted fit 15

APPENDIX A ERROR MESSAGES

CONDITION NOT MET THE 1ST TIME!

XTX_(2) IS SINGULAR!

IN CALCPW: QL IS SINGULAR

DAL IS SINGULAR

MATRIX L*L IS SINGULAR !!

QL IS SINGULAR

MATRIX (X'X) IS SINGULAR

TOO FEW VARYING PARAMETERS

XTX IS SINGULAR

MATRIX Q IS SINGULAR

[TRACE3] ERROR COMPUTING INVERSE OF L'

MATRIX Q*Q' IS SINGULAR !

NOT ENOUGH MEMORY TO RUN FLEXIFIT

The FLEXIFIT program must have at least 15k bytes of memory available for use, after FLEXIFIT is launched. (Macintosh only)

NOT ENOUGH MEMORY TO FIT DATA

The FLEXIFIT program must have at least 15k bytes of memory available in order to do a fit. (Macintosh only)

NOT ENOUGH MEMORY TO PRINT

The FLEXIFIT program must have at least 50k bytes of free memory available for use in order to print text or graphics. (Macintosh only)

ERROR READING DATA FILE

A problem occurred while reading the input data file. Most likely, an X or Y data point contains an alpha character.

MAXIMUM NUMBER OF DATA POINTS HAS BEEN EXCEEDED

The number of points allowed in an input data file is limited. Shorten the data file or reduce the number of curves.

MAXIMUM NUMBER OF CURVES HAS BEEN EXCEEDED

The number of curves in the input data file must be reduced.

APPENDIX B REPRINTS

A model-free approach to estimation of relative potency in dose-response curve analysis.

Vincenzo Guardabasso, David Rodbard, and Peter Munson , *Am. J. Physiol.* 252, E357-E364, 1987.

A versatile method for simultaneous analysis of families of curves.

V. Guardabasso, P.J. Munson, and D. Rodbard *FASEB J.* 2: 209-215; 1988.

A model-free approach to estimation of relative potency in dose-response curve analysis

VINCENZO GUARDABASSO, DAVID RODBARD, AND PETER J. MUNSON
*Laboratory of Theoretical and Physical Biology, National Institute of Child Health
and Human Development, National Institutes of Health, Bethesda, Maryland 20892*

GUARDABASSO, VINCENZO, DAVID RODBARD, AND PETER J. MUNSON. A model-free approach to estimation of relative potency in dose-response curve analysis. *Am. J. Physiol.* 252 (Endocrinol. Metab. 15): E357-E364, 1987.—We have developed a new, general approach to analysis of dose-response curves from bioassay, immunoassay (including radioimmunoassay, immunoradiuretic assay, enzyme-linked immunosorbent assay), and other experimental procedures. It provides a test for parallelism, similarity of shape, and a measure of relative potency for any set of two or more curves. The method uses a constrained smoothing spline function to estimate the curve shape, together with a nonlinear least-squares fitting technique to estimate parameters for relative potency and slope. The use of "constrained splines" permits the analysis of nonlinear dose-response curves that cannot be described by a simple model or equation such as the symmetric four-parameter logistic. A microcomputer program is used for the analysis, providing relative potencies and their SE and evaluation of goodness of fit.

data analysis; constrained curve fitting; families of curves; constrained splines; bioassay; immunoassay; computer analysis

COMPARISON of two or more dose-response curves is frequently required in studies of physiology, pharmacology, immunology, toxicology, and many other areas of biology and medicine. The ability to test whether two curves are parallel, and more generally, whether they have the same shape is essential. When the logarithm of dose is used as the abscissa, the horizontal distance between two parallel dose-response curves is the logarithm of the relative potency¹ of the two assayed compounds. The widespread use of parallel-line potency estimation (Fig. 1A) has led to very extensive development of bioassay statistics, most notably by Finney (10) and Bliss (2). These "log-linear" methods of analysis provided tests for parallelism, estimates of doses at a specified response (e.g., ED₅₀), and potency ratios (relative potencies) with their confidence limits. In the past results were computed manually or with calculators. The availability

¹ The relative potency of a test substance relative to a standard preparation is the ratio of equipotent doses of the two substances. The log of the relative potency corresponds to horizontal distances between the two curves when log(dose) are plotted. When two log(dose)-response curves are parallel, the relative potency is usually expressed as the ratio of the two ED₅₀ values (the ED₅₀ being the "effective dose" at 50% of maximal response). In case of nonparallelism the relative potency depends on the response level used; the applicability of the relative potency concept in this case has been debated extensively (3, 11).

of mainframe computers greatly facilitated the analysis and made possible the use of a larger degree of flexibility in experimental design (e.g., permitting use of unequal number of replicates or unequal spacing of dose levels) (8).

Early methods required linearity of the response curve. As such, one had to truncate the observed data to a central linear segment, or to transform data to achieve linearity. The log-linear model almost always represents an approximation: as dose is extrapolated (to 0 or infinity), the predicted response becomes plus or minus infinity, which is a physical impossibility. Rather, biological responses almost always tend to a minimal value (often 0) and to a maximal plateau value in the extremes. Hence the log(dose)-response curve must become curvilinear for small and for very large doses (Fig. 1B).

Use of a straight line to approximate a curvilinear relationship can introduce bias, especially if the mean responses of standard and unknown are different for the two curves (Fig. 1C). Such approximation can lead to a spurious indication of nonparallelism and to bias in the estimates of the relative potency and its SE (17).

A more general and useful model is the "four-parameter logistic" (Fig. 1D) (12, 26). This model provides upper and lower extremes for response and a central linear region. Most of the "parallel line" bioassay methods can be regarded as a simplification of or approximation to the logistic. A method for simultaneous analysis and comparison of several logistic curves, subject to "equality constraints" on the parameters, has been described previously (4). These constraints permit "sharing" of parameters, and enable one to set selected parameters equal to a desired constant. This approach facilitates objective testing of parallelism for dose-response curves. The constrained four-parameter logistic has enjoyed considerable popularity for bio- and immunoassay, receptor assays and related applications. However, it is not universally applicable: it cannot deal effectively with asymmetric curves (Fig. 2A), biphasic (nonmonotonic) curves (Fig. 2B), "multistep" titration curves (Fig. 2C) or even logistic curves when data are available from only a narrow region (Fig. 2D).

The problem of asymmetry in the curve has been handled by introduction of "asymmetry parameters" (22, 23, 28) in the logistic model. Fitting of these more complex models often requires more data than is available in a single assay or experiment. Accordingly, it may be

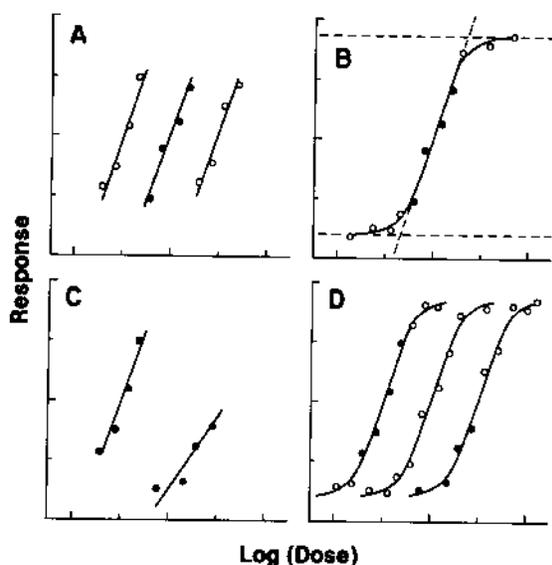


FIG. 1. A: schematic representation of a classical approach to parallel line bioassay. B: dose-response curves are almost always curvilinear, with a minimum and a maximum response. C: when 2 curves cover different ranges of levels of responses, then use of linear approximation can yield biased estimates of slope and relative potency. D: a symmetric 4-parameter logistic model is adequate for most, but not all, dose-response curve analyses.

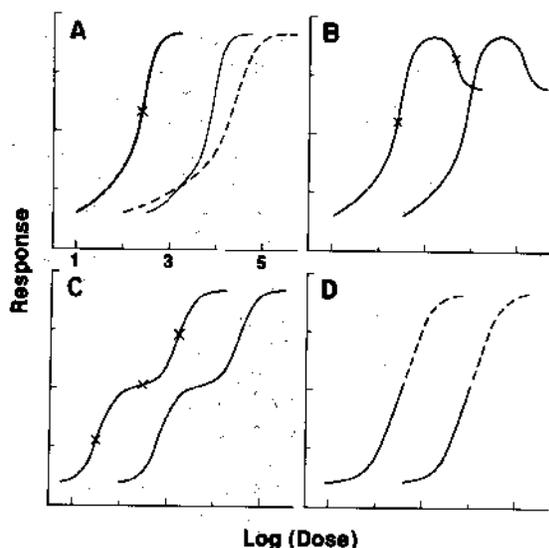


FIG. 2. A: three asymmetric sigmoidal curves: curve 2 (thin line) is identical to curve 1 (heavy line) but shifted to right by 1.5 units on a common logarithmic scale (relative potency is 0.03); curve 3 (dashed line) is not parallel to others, it covers a range twice that of curve 1. B: nonmonotonic curves. C: multistep titration curves. D: limited range of response makes it difficult or impossible to use 4-parameter logistic, because upper plateau is unknown or ill determined. Locations of inflection points are indicated by \times .

necessary to constrain the asymmetry parameters to predefined values estimated by pooling results from several experiments.

When the molecular events underlying the response being studied are known (such as in binding studies and immunoassays), models based on the Mass-Action Law

can be used (9, 25, 29, 33). Because these models can reach a substantial level of complexity, their practical applicability is often restricted to a limited range of assays and conditions and to simulation studies.

A number of empirical methods for analysis of dose-response curves have been described in the context of immunoassays (25). In this group, spline functions represent a popular method for fitting immunoassay dose-response curves (16, 18, 19). These piecewise polynomial functions provide a computer-based analysis, involve few assumptions, and can describe nonlinear curves with relatively low computational complexity. However, several problems are associated with use of one of various types of spline function (28). These problems include the arbitrary location of knots (points of connection of adjacent polynomial segments), arbitrary selection of smoothing parameters, sensitivity to outliers, tendency to oscillations and undesired nonmonotonicity, difficulties in fitting sigmoidal curves with plateaus, and lack of proper weighting based on predicted response. Further, many implementations of these empirical methods have failed to provide statistical analyses to evaluate parallelism, goodness of fit, and to estimate SE or confidence limits.

Based on our experience with these methods, and with computer programs for bioassay (8), radioimmunoassay (25), and families of dose-response curves in general (4), it became clear that there was a need for a universal model-free method for analysis and parallelism testing of dose-response curves of arbitrarily complex shape. One should not need to transform or truncate the data to achieve linearity, be forced to restrict the analysis to a monotonic region, have to oversimplify the model leading to systematic lack of fit, nor be forced to use models of excessive mathematical complexity. However, one would like to retain the features of the constrained four-parameter logistic: tests of parallelism and similarity of shape and detailed statistical analyses (4).

When presented with the curves in Fig. 2A, one has no difficulty in intuitively determining that the three curves have similar shape; one curve is shifted to the right, while another is stretched out, having a flatter slope, compared with the others. What is needed here is the ability to extract the common curve shape or template, test for similarity of shape and parallelism, and estimate the horizontal shift between two curves with its degree of uncertainty.

We chose to implement these features by using a spline function as the common curve shape. The flexibility of the spline is controlled by imposing certain constraints, e.g., monotonicity. We provide an algorithm for simultaneous least-squares fitting of this template to all curves, with objective tests for similarity and parallelism. Similarity of shape for two or more curves is defined by the ability to superimpose the data by horizontal stretching or shifting. Parallelism between curves is evaluated by checking whether the data can be superimposed allowing only horizontal shift. When curves are parallel, relative potency is determined from the magnitude of the shift on the horizontal axis, which represents logarithm of dose.

METHODS

Definition of common curve shape. A smooth, continuous curve defined by a cubic spline function is used to describe the dose-response relationship. A cubic spline function is a sequence of cubic polynomial segments joined at selected locations (knots), where continuity conditions are satisfied for the first and second derivatives (34). We have selected a smoothing spline algorithm (24). Here a smoothness parameter, S , controls the degree of complexity of the shape of the curve. The knots are located at each data point, so that no arbitrary decisions are necessary.

An appropriate choice of S is needed to obtain a satisfactory fit of the data. Various authors have suggested empirical rules for the choice of S (30). We took an alternate approach: the overall smoothness is automatically adjusted by searching for the spline with the smallest S factor that satisfies some constraints, e.g., monotonicity or a specified number of inflection points. The monotonicity constraint overcomes the tendency of spline functions to overshoot horizontal plateaus of many dose-response curves. Most calibration or assay methods require a monotonic (increasing or decreasing) standard curve. A second constraint is obtained by restricting the number of inflection points (i.e., changes in direction of curvature) within the curve. Straight lines, parabolas, and hyperbolas have no inflections; sigmoidal curves have one inflection (Fig. 2A), while biphasic curves (Fig. 2B) generally have two. More inflection points may be needed to represent more complex shapes (Fig. 2C). By controlling the number of inflection points, we can adjust the flexibility of the shape function and the complexity of the curve.

The spline can also be subjected to weighting, to compensate for nonuniformity of variance,² or to force the curve closer to the data points in areas of special interest, e.g., in regions of high curvature. When a curve has changing curvature across the range of doses, the degree of smoothing should vary for different regions. We have accomplished this by weighting the data points in proportion to the square of the second derivative when computing the spline function.

Definition of relative potency. Each member of a family of dose-response curves can be represented by a single common shape (spline function), which has been shifted horizontally to match the data for each compound. The horizontal shift parameter, designated c , is calculated for each curve relative to the reference curve. This parameter represents the base 10 logarithm of the relative potency.

Evaluation of parallelism. We can also include provision for stretching each curve in the horizontal direction. This stretch parameter b will adjust the slope of each curve with respect to the reference curve. The b parameters can be used for testing parallelism: constraining all curves to be parallel (all the $b = 1$) should not result in a significant loss of goodness of fit.

² Weights are inversely proportional to the expected variance of the predicted response. A linear, parabolic, or power function relationship between variance of response and response level is obtained by preliminary analysis of the response-error relationship (27). Weights are adjusted after each iteration (iterative reweighting).

TABLE 1. Algorithm to fit a family of dose-response curves using parallel constrained splines (FLEXIFIT)

1. Get initial estimates for scale and shift parameters b_i and c_i . We may take $b_i = 1$ in most situations.
2. Shift and scale the data (x is \log_{10} of dose, y is response)
 - for curve 1: $x' \leftarrow x$;
 - for curve 2: $x' \leftarrow b_2(x - c_2)$;
 - for curve 3: $x' \leftarrow b_3(x - c_3)$;
 - continue likewise for remaining curves
3. Fit smoothing spline f_S (i.e., the common shape function) to the shifted data (x') for all curves combined, with the algorithm given by Reinsch (24). Adjust the smoothness parameter S until desired monotonicity and desired number of inflection points are obtained.
4. Now consider the full model: $y = f_S[b_i(x - c_i)]$. Fit the nonlinear parameters b_i, c_i for every curve i except the first using the Marquardt-Levenberg nonlinear least-squares minimization algorithm (4, 21).
5. Repeat steps 2-4 until the weighted sum of squared residuals changes less than 0.01%.

Estimation of curve shape, relative potency, and parallelism. The desired constraints on the shape (monotonicity and inflection points) are first selected. An iterative procedure (Table 1) utilizes successive approximations to refine the common spline shape and parameter estimates. Initial values are needed for the parameters c (horizontal shift, i.e., logarithm of relative potency) and b (slope factor) for each curve. Starting estimates can easily be calculated by comparing the means, medians or ranges of dose levels, or by use of log-linear or logistic models as first approximations. Parameters can be fixed at constant values, shared between curves or fit individually. For example, we obtain parallel curves by constraining several b parameters to be equal to one. Forcing both the b and c parameters to be shared provides a test of identity for a set of curves.

The method (Table 1) proceeds by alternately fitting the common spline function to the superimposed data and then determining, by nonlinear least-squares regression (4, 21), the scaling parameters b and c based on that spline function. This iterative procedure is continued until convergence is reached, i.e., when the weighted sum of squares (WSS) does not change appreciably, using a specified tolerance.

Final results include relative potencies for each curve, with estimates of their SE, and evaluation of goodness of fit (see below).

Testing for parallelism with the standard or reference curve is done by fitting the data twice, once with and once without the constraint that all b parameters are equal to unity, and comparing the two fits.

Test of similarity of shape and parallelism. Poor fit in the simultaneous analysis of a family of curves might indicate that one curve has a shape different from the others, that groups of curves of different shape are being analyzed, or that the constraints imposed on the shape of all curves are too strict or inappropriate.

Several statistical criteria can be used to evaluate goodness of fit. Each curve can be fitted individually with a spline; for each curve, values for WSS and total curvature³ are computed and compared with those ob-

³ The total curvature of a spline function is the integrated squared second derivative over the entire range of the curve. It is a measure of the amount of smoothing of the spline (24); for smaller curvature values the spline function is smoother.

tained for the same curve when fitted simultaneously with other curves using a common shape. The WSS should also be compared with independent estimates of the within-dose error variance whenever replicate measurements are available for the same dose level.

The total WSS for all curves can be compared between fits performed with different constraints on the shape and on the scaling parameters. The "extra sum of squares" principle (4, 5, 21) can be used as the basis for an F test for the comparison of different fits, utilizing an approximate estimate of the effective number of degrees of freedom.

Tests for randomness of residuals (deviations of observed from predicted responses), such as the Mean Square Successive Differences test (MSSD) and the Runs test (1), are also used. If the curves have the same shape and the constraints used (monotonicity, inflections, parallelism, and others) were appropriate, the observations should be randomly distributed above or below the fitted curve. Any significant nonrandomness detected by these tests would indicate lack of fit. In addition, tests for gaussianness of residuals and tests for outliers can be applied.

EXAMPLES

An asymmetric sigmoid curve. A radioimmunoassay (RIA) standard curve was generated (Fig. 3A) using an asymmetric logistic equation (28), with parameters chosen to simulate results from a typical radioimmunoassay. A normally distributed random error was added to the predicted y values.

Figure 3A shows a four-parameter logistic model fit to the data. Upon casual inspection, the fit appears adequate, but careful analysis of the residuals (top) indicates a nonrandom pattern and hence significant lack of fit. The small but consistent departures shown by the logistic fit can lead to significant errors in dose interpolation and potency estimation for unknown samples. Figure 3B shows a smoothing spline fit to the data, constrained to be monotonic and to have only one inflection point. The

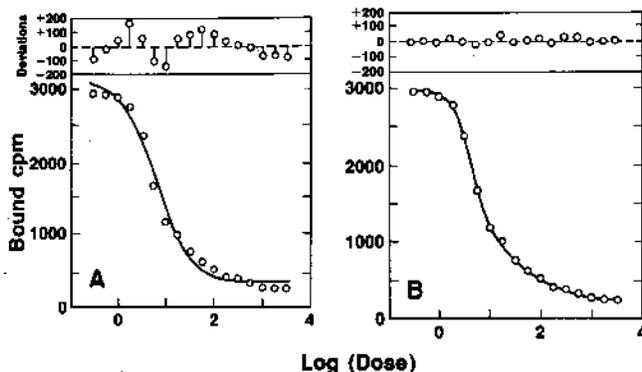


FIG. 3. Radioimmunoassay standard curve simulated by using an asymmetric logistic equation (28), with uniform simulated measurement error. A: fit of a 4-parameter symmetric logistic equation. Top shows residuals—difference of observed and predicted values. Systematic lack of fit is shown by nonrandom pattern of residuals. B: fit of a constrained spline function to same data; monotonicity and 1 inflection point were specified. Analysis of residuals (top) shows deviations to be small and random.

constrained spline function fits the data well with a random pattern of residuals and a substantially smaller root mean square (RMS) error.

Bioassay: parallelism and relative potencies. Figure 4A shows the application of this technique to some dose-response curves from a bioassay of four different aminoglycoside antibiotics (13). The response is the inhibition of contractility in a preparation of longitudinal muscle-myenteric plexus from guinea pig ileum. The experimental curves were analyzed simultaneously, constraining the common spline function to be monotonic and to have at most one inflection point. Constraining the curves to be parallel (by setting the slope factors b for all curves equal to unity) resulted in no indication of lack of fit, so we infer that the curves are parallel. Potencies relative to the most potent compound (streptomycin) were estimated as 0.74 ± 0.04 (dihydrostreptomycin), 0.61 ± 0.03 (kanamycin), and 0.27 ± 0.01 (gentamycin), respectively. Figure 4B shows the superimposition of the data from all curves after shifting the curves horizontally to the left according to their relative potencies.

Parallelism of the standard and unknown dilution curves. Figure 5A shows data from an ELISA (enzyme-

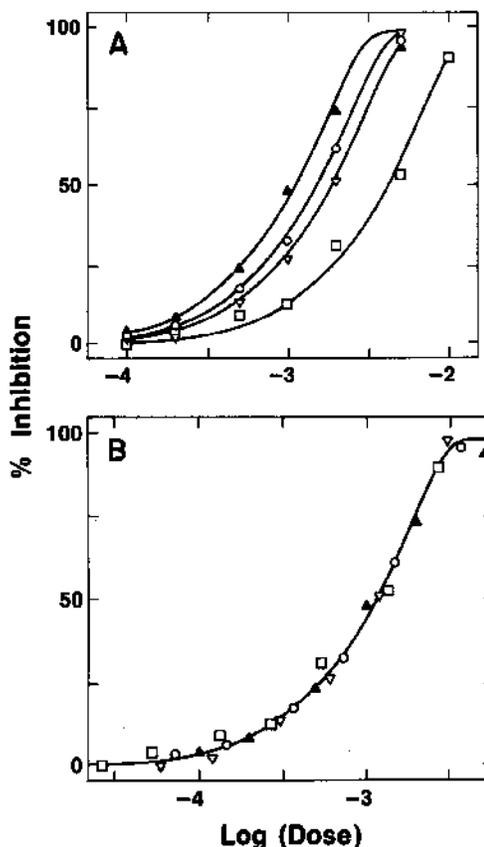


FIG. 4. Inhibition of contractility of a preparation of longitudinal muscle-myenteric plexus from guinea pig ileum, by 4 different aminoglycoside antibiotics (13). A: dose-response curves for 1) streptomycin (solid triangles); 2) dihydrostreptomycin (circles); 3) kanamycin (open triangles); 4) gentamycin (squares). Simultaneous fit, constrained for parallelism and monotonicity, with one inflection point. B: same data superimposed after horizontal shift of curves 2, 3, and 4 to left according to best estimates of their relative potencies.

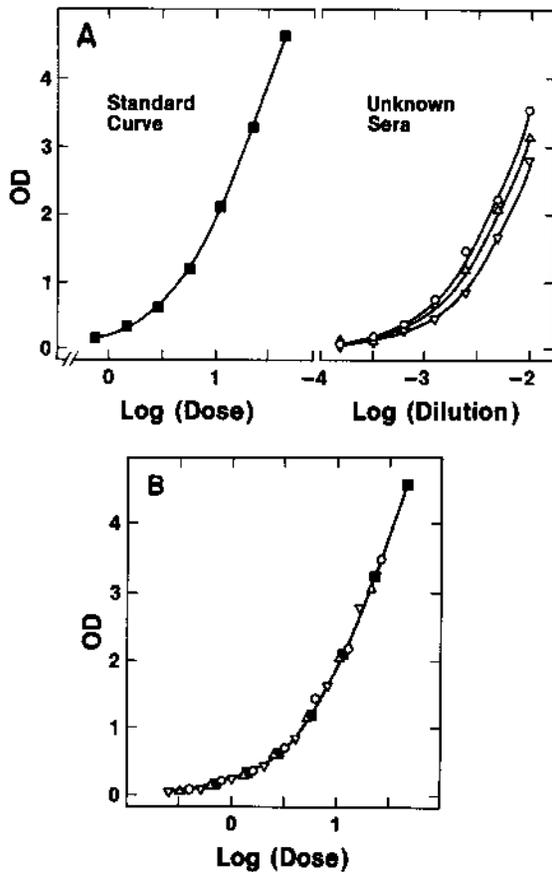


FIG. 5. ELISA assay for bacterial outer membranes (31). Ordinate: optical density adjusted to 100 min. *A*: left, standard curves; right, curves from dilutions of 3 serums of unknown potency. A spline function template was fit to all 3 curves simultaneously, constraining curves to be parallel and with no inflection points. *B*: curves superimposed to show fit of common spline function, after optimal horizontal shifting using estimates of c parameters.

TABLE 2. Analysis of standard and dilution curves for three unknowns, for data of Fig. 5

Fit	Constraint	Sum of Squares ($\times 10^{-2}$)	Randomness of Residuals	
			MSSD	Sign of residuals + -
1	Common shape, nonparallel	2.73	NS	16 12
2	Common shape, parallel curves	5.84	$P < 0.05$	19 9
3	Common shape, 1 curve nonparallel, others parallel	3.17	NS	19 9

MSSD, mean square successive difference criterion (1); NS, not significant at the $P = 0.05$ level.

linked immunosorbent assay) for an antibody to bacterial outer membranes (31). A standard curve (left) was analyzed simultaneously with three curves (right) obtained from the dilution of three test sera (unknowns), assuming parallelism. Fig. 5B shows the data superimposed on the common spline function. The common spline function

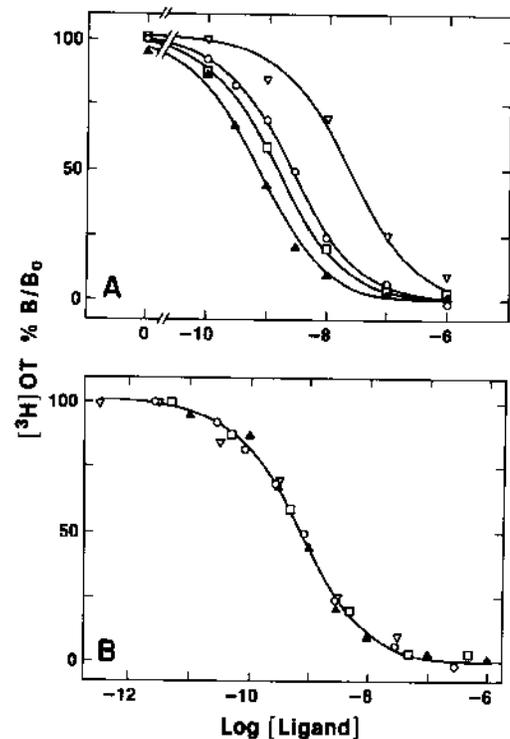


FIG. 6. Displacement of $[^3H]$ oxytocin from porcine epididymis membrane receptors (15), by oxytocin (solid triangles), arginine vasopressin (squares), $[Thr^4,Gly^7]oxytocin$ (circles), and $d(CH_2)_5Tyr(Me)arginine$ vasopressin (open triangles). *A*: fit of a common spline template, allowing for horizontal shift and constraining curves to be parallel. *B*: data superimposed after horizontal shift corresponding to relative potency.

TABLE 3. Analysis of radioreceptor assay data of Fig. 6

Fit	Constraint	Sum of Squares	
		Present method	Four-parameter logistic method*
1	Common shape	103	126
2	Common shape, parallel curves	198	179

Curve	Potency Relative to Curve 1 (in fit 2)	
	Present method	Four-parameter logistic method*
2	0.452 ± 0.059	0.461 ± 0.069
3	0.267 ± 0.026	0.274 ± 0.034
4	0.028 ± 0.004	0.028 ± 0.004

* By use of program ALLFIT (4).

was chosen to be monotonic, without an inflection point. Although a single shape function can be used without appreciable lack of fit, imposition of the constraint of parallelism resulted in lack of fit, based on the MSSD criterion (Table 2, fit 2). By allowing just one curve to have a different slope factor, goodness of fit became satisfactory (Table 2, fit 3).

These data could not be readily analyzed using a four-parameter logistic model because the range of concentrations was insufficient to permit observation of the full sigmoidal shape of the curve. One could set an arbitrary upper plateau, high above the curve, and constrain it to be constant. However, this may or may not provide a

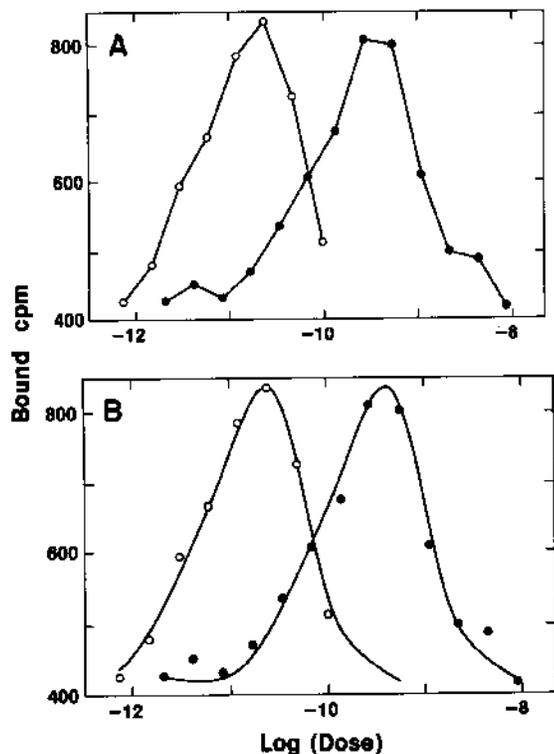


FIG. 7. Cooperative immunoassay dose-response curves (7) for human chorionic gonadotropin (hCG, open circles), and for hCG β -subunit (solid circles). Ordinate: counts of bound ^{125}I -hCG. A: point-to-point interpolation of data. B: common spline function fit to the 2 curves with best estimate of relative potency.

satisfactory fit for the observed data and introduces greater complexity, with very ill-defined and highly interacting parameters for the ED_{50} and the slope. Many laboratories might resort to use of point-to-point interpolation. Alternatively one could try simple curvilinear (quadratic, exponential, or power function) models on an ad hoc basis. A logit transformation, calculated using an arbitrary guess for the maximal response, might lead to biased estimation. Because the true maximal response is not defined by the data, an ED_{50} for the standard curve cannot be estimated.

Radioreceptor assay. Several dose-response curves were generated by measuring the inhibition of binding of [^3H]oxytocin ([^3H]OT) to porcine epididymis (15). Oxytocin itself and various analogues were used as competing unlabeled ligands. The response shown is the binding of [^3H]OT, expressed as $\%B/B_0$, i.e., normalized to 100% in absence of unlabeled ligand. Simultaneous analysis of the curves (Fig. 6A) allows the fitting of a common sigmoidal shape (Fig. 6B), subject to constraints of monotonicity and a single inflection point. Because the curves appear symmetric, sigmoidal, and extend over a wide enough range, it was possible to compare the results from the present constrained spline approach with those from a four-parameter logistic fit (4) obtained by constraining all curves to share parameters a (minimal response), d (maximal response), and b (slope factor) (Table 3). Similar results were obtained using both methods for goodness of fit, relative potencies, and their standard errors.

Nonmonotonic curves. Calibration or standard dose-response curves obtained from cooperative immunoassays (CIA) (6) are often nonmonotonic. Response (counts of bound labeled antigen) increases and then decreases with increasing concentration of antigen, apparently due to formation of circular complexes. Mathematical models for these assays have been described (20). However, these models require many assumptions, involve many parameters, and their estimation is complex (7). Often all that is required by the investigator is an estimate of relative potency between two curves, after a test for parallelism.

Figure 7A shows two CIA curves: on the left is a dilution curve for human chorionic gonadotropin (hCG), on the right a dilution curve for the β -subunit of hCG. Two inflection points were needed to represent the bell-shaped curves. Figure 7B shows the same spline function drawn on the two sets of data in the original coordinates. The spline fit was satisfactory; the relative potency of the hCG β -subunit with respect to hCG is estimated as 0.058 ± 0.002 . This was in excellent agreement with the results obtained using the four-parameter logistic model, to describe upstrokes or downstrokes separately (data not shown).

DISCUSSION

We have developed a simple, generally applicable, empirical method to obtain objective descriptions of families of dose-response curves. This versatile model-free approach can be applied even when the mathematical model (equation) is unknown, when data are insufficient to fit the "true" underlying mathematical model or when a simplified model shows consistent lack of fit. Spline functions provide the desirable features of continuity and smoothness, and in our method they can be constrained by a priori knowledge of the expected degree of complexity of shape for a particular application.

The constrained spline function is computed from a large set of data points, obtained by pooling data from a family of curves. The usefulness of simultaneous fitting of multiple curves has long been advocated. Indeed, the use of "analysis of covariance" to pool information from several curves regarding a common slope and test the hypothesis of parallelism, is a basic tenet of bioassay statistics applied to the log-linear model (2, 10). Lawton et al. (14) describe a closely related method for analysis of population of curves by scaling all curves in reference to a common shape function. For this shape function they used a polygonal curve, constituted of a series of linear segments connected at selected locations. However, they did not address the problems of testing goodness of fit, similarity of shape and parallelism, or estimating relative potency and its SE or 95% confidence limits. Waud (32) suggests simultaneous fitting of several sets of pharmacological data. De Lean et al. (4) describe the advantages of simultaneous fitting of families of dose-response curves using a general four-parameter logistic model. Munson and Rodbard (21) use simultaneous fitting of multiple binding curves to estimate parameters and select models for characterization of complex binding systems. The present method for simultaneous fitting, employing a single common smooth function of

arbitrary shape, provides a general method for analysis of families of dose-response curves.

The imposition of constraints on the common shape solves many of the problems previously associated with the use of spline functions for immunoassay data analysis (28), including undesired nonmonotonicity, oscillations, and difficulty in fitting curves with plateaus or asymptotes. Iterative weighting is used to compensate for any systematic nonuniformity of variance of response. We avoid the need for arbitrary or subjective decisions to specify location of knots or the degree of smoothing. Choice of proper constraints (monotonicity, inflection points, weighting proportional to curvature) is based on the shape of the data curves. Constrained fitting of a common spline template can be applied to data sets for which no explicit mathematical model is available or practical. The present method can also be applied in cases where a more traditional "modeling" approach is feasible and yields comparable results. Although the method requires fairly extensive computations, it is practical even when performed using a microcomputer.

A number of ancillary features can be added: For curves with maxima and minima, the ED_{50} can be computed. Parameters for vertical as well as horizontal scaling can be introduced (V. Guardabasso et al., unpublished observations). Vertical scaling can account for both shifting and stretching in the y -axis, providing further generality and flexibility of the present approach.

The method described here for fitting of families of curves with a common shape provides a general approach for computer fitting of dose-response data. In principle, the method should be almost universally applicable, provided only that the response represents a smooth continuous function of dose. The present method has the virtues of simplicity and flexibility as in other empirical approaches, yet it can provide objectivity, statistical evaluation, and hypothesis testing comparable with those obtained by mathematical modeling.

The authors thank Drs. Paul Ehrlich, Carl Frasch, Daniel Goldstein, Mario Maggi, William Moyle, and Thomas Ropchak for providing illustrative data.

Received 5 May 1986; accepted in final form 23 October 1986.

REFERENCES

- BENNETT, C. A., AND N. L. FRANKLIN. *Statistical Analysis in Chemistry and the Chemical Industry*. New York: Wiley, 1954.
- BLISS, C. I. *The Statistics of Bioassay*. New York: Academic, 1952.
- CORNFIELD, J. Comparative bioassays and the role of parallelism. *J. Pharmacol. Exp. Ther.* 144: 143-149, 1964.
- DE LEAN, A., P. J. MUNSON, AND D. ROBBARD. Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves. *Am. J. Physiol.* 235 (Endocrinol. Metab. Gastrointest. Physiol. 4): E97-E102, 1978.
- DRAPER, N. R., AND H. SMITH. *Applied Regression Analysis*. New York: Wiley, 1966.
- EHRlich, P. H., AND W. R. MOYLE. Cooperative immunoassays: ultrasensitive assays with mixed monoclonal antibodies. *Science* 221: 279-281, 1983.
- EHRlich, P. H., AND W. R. MOYLE. Specificity considerations in cooperative immunoassays. *Clin. Chem.* 30: 1523-1532, 1984.
- FADEN, V. B., AND D. ROBBARD. *BIOPROG: Bioassay Data Processing*, Bethesda, MD: NIH, NICHD, 1975.
- FELDMAN, H., D. ROBBARD, AND D. LEVINE. Mathematical theory of cross-reactive radioimmunoassay and ligand-binding systems at equilibrium. *Anal. Biochem.* 45: 530-556, 1972.
- FINNEY, D. J. *Statistical Method in Biological Assay* (2nd ed.). London: Griffin, 1964.
- FINNEY, D. J. The meaning of bioassay. *Biometrics* 21: 785-798, 1965.
- FINNEY, D. J. Radioligand assay. *Biometrics* 32: 721-740, 1976.
- GOLDSTEIN, D. J., T. G. ROPCHACK, H. R. KEISER, A. ARGIOLAS, AND J. J. PISANO. Bradykinin and the gut cholinergic system. In: *Kinin 84*, edited by L. Greenbaum and H. Margolius. New York: Plenum. In press.
- LAWTON, W. H., E. A. SYLVESTRE, AND M. S. MAGGIO. Self-modeling nonlinear regression. *Technometrics* 14: 513-532, 1972.
- MAGGI, M., S. MALOZOWSKI, S. KASSIS, V. GUARDABASSO, AND D. ROBBARD. Identification and characterization of two classes of receptors for oxytocin and vasopressin in porcine tunica albuginea, epididymis, and vas deferens. *Endocrinology*. In press.
- MALAN, P. G., M. G. COX, E. M. R. LONG, AND R. P. EKINS. Curve fitting to radioimmunoassay standard curves: spline and multiple binding-site models. *Ann. Clin. Biochem.* 15: 123-135, 1978.
- MANTEL, N., AND A. G. HILGAR. Effects of nonlinearity on subjective and objective analyses of hormone assay data. *Cancer Chemother. Rep.* 11: 61-68, 1961.
- MARSCHNER, I., F. ERHARDT, AND P. C. SCRIBA. Calculation of the radioimmunoassay standard curve by "spline function." In: *Radioimmunoassay and Related Procedures in Medicine*. Vienna: Int. Atomic Energy Agency, 1974, vol. 1, p. 111-122.
- MOSLEY, J., AND B. R. BEVAN. Spline function analysis applied to an human placental lactogen assay. *Ann. Clin. Biochem.* 14: 16-21, 1977.
- MOYLE, W. R., C. LIN, R. L. CORSON, AND P. H. EHRlich. Quantitative explanation for increased affinity shown by mixtures of monoclonal antibodies: importance of a circular complex. *Mol. Immunol.* 20: 439-452, 1983.
- MUNSON, P. J., AND D. ROBBARD. LIGAND: a versatile computerized approach for characterization of ligand-binding systems. *Anal. Biochem.* 107: 220-239, 1980.
- PRENTICE, R. L. A generalization of the Probit and Logit methods for dose-response curves. *Biometrics* 32: 761-768, 1976.
- RAAB, G. M. Comparison of a logistic and a mass-action curve for radioimmunoassay. *Clin. Chem.* 29: 1757-1761, 1983.
- REINSCH, C. H. Smoothing by spline functions. *Numer. Math.* 10: 177-183, 1967.
- ROBBARD, D. Data processing for radioimmunoassay: an overview. In: *Clinical Immunochemistry: Chemical and Cellular Bases and Applications in Disease*, edited by S. Natelson, A. J. Pesce, and A. A. Dietz. Washington, DC: Am. Assoc. for Clin. Chem., 1978, p. 477-494.
- ROBBARD, D., AND D. M. HUTT. Statistical analysis of radioimmunoassays and immunoradiometric (labeled antibody) assays: a generalized, weighted, iterative least-squares method for logistic curve fitting. In: *Radioimmunoassay and Related Procedures in Medicine*. Vienna: Int. Atomic Energy Agency, 1974, vol. 1, p. 165-192.
- ROBBARD, D., R. H. LENOX, H. L. WRAY, AND D. RAMSETH. Statistical characterization of the random errors in the radioimmunoassay dose-response variable. *Clin. Chem.* 22: 350-358, 1976.
- ROBBARD, D., P. J. MUNSON, AND A. DELEAN. Improved curve-fitting, parallelism testing, characterization of sensitivity and specificity, validation, and optimization for radioligand assays. In: *Radioimmunoassay and Related Procedures in Medicine*. Vienna: Int. Atomic Energy Agency, 1978, vol. 1, p. 469-504.
- ROBBARD, D., AND R. L. TACEY. Radioimmunoassay dose interpolation based on the mass action law with antibody heterogeneity. *Anal. Biochem.* 90: 13-21, 1978.
- WAHBA, G., AND S. WOLD. A completely automatic french curve:

- fitting spline functions by cross validation. *Commun. Stat.* 4: 1-17, 1975.
31. WANG, L. Y., AND C. E. FRASCH. Development of a *Neisseria meningitidis* group B serotype 2b protein vaccine and evaluation in a mouse model. *Infect Immun.* 46: 408-411, 1984.
 32. WAUD, D. R. Analysis of dose-response curves. In: *Methods in Pharmacology*, edited by E. E. Daniel and D. M. Paton. New York: Plenum, 1975, vol. 3, chapt. 27.
 33. WILKINS, T. A., D. C. CHADNEY, J. BRYANT, S. H. PALMSTROM, AND R. L. WINDER. Nonlinear least-squares curve-fitting of a simple theoretical model to radioimmunoassay dose-response data using a mini-computer. In: *Radioimmunoassay and Related Procedures in Medicine*. Vienna: Int. Atomic Energy Agency, 1978, vol. 1, p. 399-423.
 34. WOLD, S. Spline function in data analysis. *Technometrics* 16: 1-11, 1974.

A versatile method for simultaneous analysis of families of curves

V. GUARDABASSO,¹ P. J. MUNSON, AND D. RODBARD²

Laboratory of Theoretical and Physical Biology, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

ABSTRACT

We have developed a versatile new approach to the simultaneous analysis of families of curves, which combines the simplicity of empirical methods with several of the advantages of mathematical modeling, including objective comparison of curves and statistical hypothesis testing. The method uses weighted smoothing cubic splines; the degree of smoothing is adjusted automatically to satisfy constraints on curve shape (monotonicity, number of inflection points). By simultaneous analysis of a family of curves, one can extract the shape common to all the curves. Up to four linear scaling parameters are used to match the shape to each curve, and to provide optimal superimposition of the several curves. By applying constraints to these scaling factors, one can test a variety of hypotheses concerning comparisons of curves (e.g., identity, parallelism, or similarity of shape of two or more curves), and thus evaluate the effects of experimental manipulation. By optimal pooling of data one can avoid the need for arbitrary selection of a typical experiment, and can detect subtle but reproducible effects that might otherwise be overlooked. This approach can facilitate the development of an appropriate model. The method has been implemented in a Turbo-Pascal program for IBM-PC compatible microcomputers, and in FORTRAN-77 for the DEC-10 mainframe, and has been utilized successfully in a wide variety of applications. — GUARDABASSO, V.; MUNSON, P. J.; RODBARD, D. A versatile method for simultaneous analysis of families of curves. *FASEB J.* 2: 209-215; 1988.

Key Words: graphic and statistical analysis • microcomputer programs • mathematical modeling • nonlinear least-squares curve fitting • hypothesis testing

ANALYSIS OF SCIENTIFIC DATA MAY be regarded as belonging to one of two schools: empirical or model-based. Each has its advantages and limitations (Table 1). We would like to be able to extend the objectivity of the model-based methods to cases where no mathematical formulation is known and when the available data do not permit derivation of such a model. Further, we would like

to combine the information derived from multiple curves, e.g., families of curves representing multiple subjects, treatments, conditions, experiments, or assays.

A recent review of curve-fitting procedures (1) has noted that empirical methods such as polynomial regression and interpolating cubic splines are "not frequently used for data analysis in biology" and are "not often useful." In this paper, we develop an enhancement of spline-based methods that can indeed be an important, general analysis technique for many applications in experimental biology and medicine, especially those dealing with several curves and experiments. Although analysis of a single curve may fail to define adequately the curve shape, simultaneous analysis of multiple curves may provide excellent definition of the common shape or template. Differences between curves may then be characterized by horizontal and/or vertical shifting and scaling. In common practice, normalization or scaling of data is based on one or two observed points, e.g., baseline (0%) and/or maximal (100%) responses. However, it is preferable to estimate the scaling factors based on consideration of all of the data, which becomes possible once the underlying shape function or template is determined.

Lawton et al. (2) have previously proposed a method for self-modeling nonlinear regression, which achieves many of these objectives. They used a polygonal shape (i.e., a first-order spline, a continuous series of linear segments joined at "knots") to describe curves of arbitrary shape, combined with least-squares curve fitting to estimate scaling parameters. However, this approach has been used in only a limited number of studies (e.g., 3-7). With these few exceptions, the statistical, biomedical, and general scientific community has not appreciated the potential widespread applicability and importance of this combination of empirical and modeling approaches.

We report here the development of a series of modifications and refinements to the approach of Lawton et al. (2), which we believe will make the approach considerably more useful:

¹Present address: Istituto di Ricerche Farmacologiche "Mario Negri," Milan, Italy.

²To whom reprint requests should be addressed.

- 1) Use of smoothing cubic splines (8) (instead of linear segments)
- 2) Elimination of the need for the user to specify the location of the knots (junctions between segments) in the curves
- 3) Automatic estimation of an appropriate smoothing factor, based on the need to satisfy constraints regarding monotonicity or complexity of curve shape (number of inflection points)
- 4) Weighting to compensate for nonuniformity of variance (9)
- 5) Ability to fit multiple curves simultaneously, permitting the use of shared parameters for multiple curves (10, 11)
- 6) Implementation in a practical program for use both on mainframes and with the IBM-PC class of microcomputers

We have previously reported the application of this methodology to analysis of data arising in bioassays, immunoassays, and receptor assays (12, 13). In that application, only horizontal shifting or scaling of the independent variable is ordinarily required. We now consider the use of this approach in the more general context, providing the options of both horizontal and vertical shifting and/or stretching, in any combination (Fig. 1).

DESCRIPTION OF THE METHOD

This method consists of two steps: 1) determination of the common shape, 2) estimation of optimal vertical and/or horizontal scaling factors to superimpose the curves. These steps are repeated iteratively until convergence is achieved (Fig. 2; Appendix).

TABLE 1. Approaches to data analysis

Empirical school	Model-based school
Examples	Examples
Manual: visual curve fitting	Manual: graphic fit of straight line (e.g., Scatchard plot, semilog plot)
Computer: Point-to-point interpolation Polynomials Splines (interpolating or smoothing)	Computer: Mass-action law models Four-parameter logistic Multicomponent exponential decay Compartmental pharmacokinetic models
Advantages	Advantages
Simplicity Versatility Arbitrary shape No assumption of ideal model	Physicochemical parameters Objective goodness-of-fit tests Standard errors for parameters Objective hypothesis testing
Disadvantages	Disadvantages
Subjective No statistical information regarding precision No hypothesis testing No quantitative description of the data	Requires mathematical expressions May fail to fit data Subject to numerical difficulties

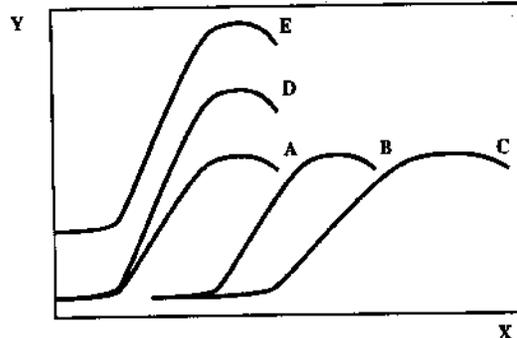


Figure 1. Schematic representation of a family of curves. Curve B is parallel to curve A, having undergone a horizontal shift. Curve C is related to A by a horizontal shift and stretch. Curve D corresponds to A with a vertical stretch (change in maximal response), whereas curve E has both a vertical shift and stretch. These curves are nonmonotonic and have one inflection point. Numerous hypotheses can be tested in terms of the four scaling factors (two horizontal, two vertical).

Step 1: determination of the common shape

Curves of arbitrary shape can be approximated by straight line segments joining the data points (Fig. 3A). However, this results in abrupt changes of slope or first derivative, in contrast to the behavior of most physical and biological systems. Curves can be better approximated by polynomials, e.g.,

$$y = k_0 + k_1x + k_2x^2 + k_3x^3 + \dots \quad (1)$$

Increasing the degree or order of the polynomial provides increasing flexibility. However, polynomials have two undesirable properties: they may oscillate between points and can show inappropriate behavior as one extrapolates beyond the range of the data.

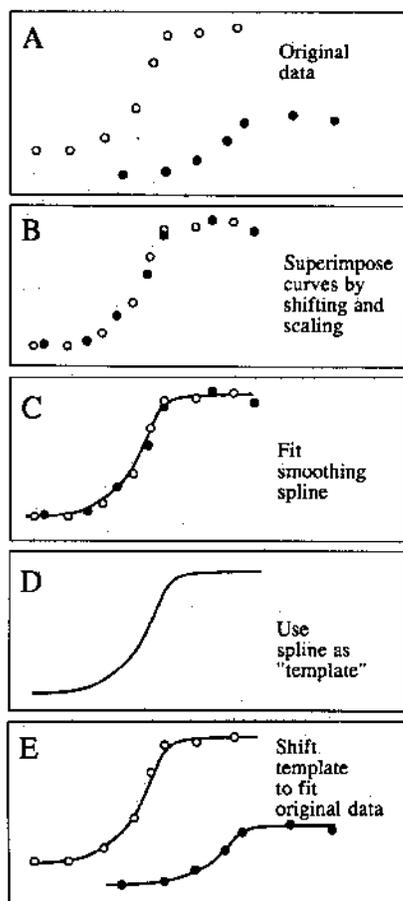


Figure 2. Schematic representation of algorithm to fit families of curves of arbitrary shape. For simplicity, only two curves are shown. *A)* Data points. *B)* The data are superimposed by vertical scaling (shifting and stretching) and horizontal shifting of the second curve relative to the first. *C)* A smoothing spline is fit to the combined data, satisfying the user-specified constraints of monotonicity and/or number of inflections, and with the appropriate weighting. *D)* The shape of the template is now fixed. *E)* The position of the template is scaled vertically and horizontally to obtain the best fit to the original data. Then the data are rescaled to obtain a new, improved superimposition (return to *B*). The entire cycle is repeated until no further changes are observed, i.e., until convergence.

Rather than progressively increasing the degree of the polynomial, one can use cubic polynomials to describe local segments of the curve, which can be joined together smoothly at knots. This is accomplished by requiring the first two derivatives of the curve to be continuous. If a knot is placed at each data point, the result is a cubic interpolating spline (Fig. 3*B*), which is the smoothest curve that can be drawn exactly through the points, where smoothness is defined as the reciprocal of the integral of the square of the second derivative, $S = 1/\int (y'')^2 dx$ (8). Using fewer knots than the number of data points, placed at locations selected by the user, one obtains least-squares splines (14): the curve will now pass close to, but not necessarily exactly through, each point. However, the results depend on the number and location of the knots, which introduces an element of subjectivity and arbitrariness.

The smoothing spline approach (Fig. 3*C*) combines the virtue of placing a knot at each data point, with the property that the curve should pass near—but not necessarily exactly through—each point in the presence of experimental errors (8, 13). Here, one must specify a smoothing factor, S , to compromise between smoothness and average nearness to each point. The smoothness factor could be determined if one has knowledge of the magnitude of random measurement error in the data (8) or by a cross-validation approach (15). Alternatively, the user may indicate the expected degree of complexity of the curve, e.g., whether or not the curve is monotonic (always increasing or decreasing), and the maximum number of inflection points permitted. These simple specifications are often adequate to define an acceptable level of smoothing. Our program will then automatically adjust the smoothing factor S until these conditions (constraints) are satisfied.

Unweighted least-squares curve fitting can give inappropriate results if the magnitude of random measurement error varies systematically along the length of the curve (e.g., constant relative error $\sigma_y^2 \propto y^2$, or the more general cases, e.g., $\sigma_y^2 \propto y^J$). To compensate for systematic non-uniformity of variance, one can use transformations (log, square root), or preferably utilize weighting, such that each point receives weight proportional to the informa-

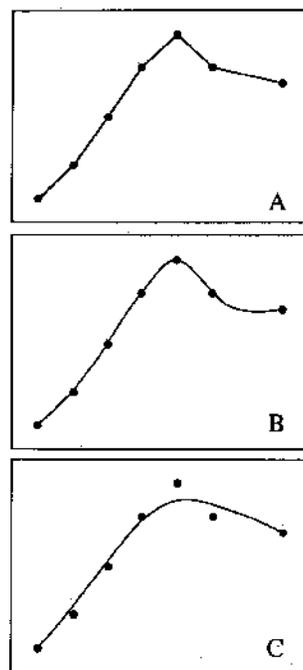


Figure 3. Three methods for empirical curve fitting. *A)* Point-to-point interpolation. Alternatively, one can use a series of continuous linear segments, joined at a fewer number of locations (knots) than number of observations. *B)* Interpolating cubic splines are the smoothest curves that will pass exactly through each point (8, 23). *C)* Smoothing spline: this method shares properties with both the interpolating and least-squares spline. A knot is placed at the x value corresponding to each point. One uses a smoothing factor to select a compromise between closeness to each observation and smoothness, by minimizing the weighted sum-of-squares of deviations added to a measure of roughness (see ref 8).

tion that it carries (9, 16). Optionally, iteratively reweighting the points in proportion to the square of the second derivative appears to make the function follow the data more closely in regions of high curvature.

If a least-squares method of curve fitting is deemed inappropriate (e.g., owing to severe nonnormality of the error distribution, frequent outliers), one could in principle use robust methods of curve fitting or smoothing, e.g., L_1 norms (17) or LOWESS (18).

Step 2: vertical and horizontal scale factors

We wish to obtain the optimal superimposition of the entire family of curves by means of horizontal and/or vertical shifting and/or stretching. We provide the use of up to four scaling parameters, corresponding to horizontal shift, horizontal stretch, vertical shift, or vertical stretch (Fig. 1). Because some of these parameters may be unnecessary, redundant, or nonidentifiable, they can be turned off by constraining them to be unity or to another constant.

If two curves are identical, then they must have identical values for all four scaling parameters. If they are parallel, then they must have the same horizontal stretch factor. Two curves can be regarded as congruent, i.e., similar in shape, if it is possible to superimpose them with any combination of the four scaling factors.

We utilize nonlinear least-squares curve fitting (19, 20), iteratively reweighted according to the variance model (9), to estimate the scaling parameters. Selected parameters can be set equal to unity (or other constants), whereas selected sets of parameters can be shared (10, 11), corresponding to similar upper and lower plateaus, similar slope (parallelism), or identity. Again, in principle, L_1 norms or other robust methods could be substituted for the least-squares method (17).

Goodness-of-fit is evaluated by a number of conventional criteria, including (weighted) sum-of-squares of deviations (SS), analysis of residuals for Gaussianness and randomness (using the runs-test and mean squared successive difference tests), and by plots of residuals and the histogram of residuals (19, 21). The effective number of degrees of freedom (df) for a smoothing spline is an interesting question. For an interpolating spline, the number of df is zero. For a smoothing spline,

it depends on the degree of smoothing. When the degree of smoothing is large, the spline approaches a straight line, which has two parameters. Thus, df approaches its upper limit of $N - P - 2$, where N is the total number of points and P is the number of scaling parameters that have been fit. Using this approximate (maximal) number of df, one can convert the weighted SS to a mean square (MS) and root mean square (RMS) error, and obtain approximate, minimal estimates of the asymptotic standard errors of parameters. Monte-Carlo, jackknife, and bootstrap methods (22) can be used to provide improved estimates of effective df, RMS, and standard errors of parameters in particular applications.

IMPLEMENTATION

The method has been implemented in computer programs in Turbo-Pascal and FORTRAN 77 for the IBM-PC class of microcomputer. A floating-point coprocessor chip is highly desirable to enhance execution speed. Table 2 summarizes the nature of the input and output of program FLEXIFIT.

APPLICATION TO HYPOTHESIS TESTING

One can test several hypotheses (e.g., identity or parallelism of two or more curves) as follows: 1) formulate the hypothesis in terms of a set of constraints on shift and scale parameters, e.g., $b_1 = b_2$; 2) fit the curves without these constraints (Fig. 4B); 3) fit the curves a second time subject to these constraints (Fig. 4A). Then, compare the sum-of-squares for the two fits, using an F test based on the extra sum-of-squares principle (19). This evaluates whether the use of additional fitted parameters in the unconstrained fit has resulted in a statistically significant reduction in the sum-of-squares. This test assumes that the df for the residual mean square have been estimated appropriately. One can also compare various fits (corresponding to various models or hypotheses) in terms of the randomness of the residuals (21).

TABLE 2. Input and output of program FLEXIFIT

Input	Output
No. of curves	Common curve template
Curves: x, y pairs	SS, MS, RMS, approximate df
Coefficients of variance model (e.g., constant σ , constant %CV, or σ as function of y)	Scaling parameters and their approximate standard errors
Monotonicity (yes/no)	Graphs of curves in original coordinate system
Maximum number of inflection points	Graph of superimposed curves (after scaling)
$(y'')^2$ weighting (yes/no)	Tests of residuals:
Constant parameters	Randomness (mean square successive difference, MSSD, and runs tests) (see ref 20)
Shared parameters	Gaussianness
	Outliers
	Graph of standardized residuals
	F test to compare alternative fits, models, constraints

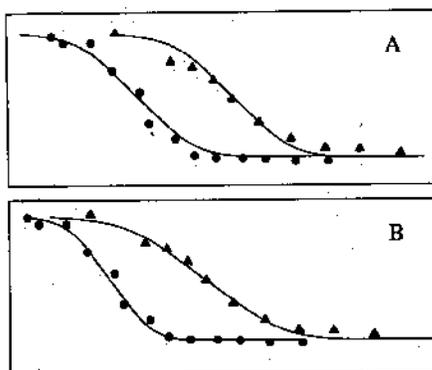


Figure 4. Use of FLEXIFIT for hypothesis testing. *A*) Two curves fit subject to the constraint of parallelism ($b_1 = b_2$). A good fit is obtained, but careful analysis of residuals suggests a subtle lack of fit. *B*) Repeat fit, allowing for horizontal stretch as well as shift, results in a major improvement in goodness-of-fit, viewed in terms of magnitude (sum-of-squares, MS, RMS) or randomness (MSSD test) of residuals. One can test numerous other hypotheses by formulating them in terms of constraints involving one or more of the scaling parameters.

DISCUSSION

Families of curves are commonly obtained from the observation of multiple experiments, assays, treatments, or subjects. We provide a versatile method for simultaneous analysis of families of curves when no a priori knowledge is available regarding the shape of the curves. This method is broadly applicable. We have previously illustrated its use in the estimation of relative potencies in bioassay, immunoassay, and similar procedures (13). This is ordinarily a simplified case, where only horizontal scaling parameters may be necessary.

The availability of vertical scaling parameters permits handling of cases where the curves have different baselines or different maximal response (e.g., curves A, D, and E in Fig. 1). In these cases the present method is superior to the use of a normalized response variable (e.g., percentage of control, percentage of maximal response, percentage between minimal and maximal response), because it uses information derived from all of the points on the curves to estimate scaling, unlike approaches based on just one or two measured values for control, minimal, or maximal response. Thus it is possible to properly scale the experimental curves even if a measure of these 0 and 100% values is missing owing to accident or experimental design.

This model-free analysis of data provides many of the advantages of mathematical modeling in cases where the experimentalist cannot supply a mathematical formulation of a model (Table 1). It facilitates pooling of results from several experimental curves, even when these involve different experimental designs (different sets of values for the independent variable). This should permit display of data from many or all of the experiments in a series, rather than the common practice wherein the experimentalist must select a typical or representative experiment for illustration. This

selection process rarely involves a random sample: often the best experiment is presented, losing information about between-experiment variability. By isolating extraneous sources of variation, as with the use of analysis of variance or analysis of covariance, the present method provides signal averaging of pooled data, and retains objectivity, allowing one to display all of the data simultaneously.

Resourceful use of constraints permits testing of many hypotheses. By selecting various combinations of scaling parameters, by excluding unwanted scaling possibilities from the model, by constraining some of the parameters to be shared between subgroups of curves or to be fixed at constant values, the user can interactively test hypotheses regarding parallelism, similarity of shape, or other features in common. Subsets of curves significantly different in shape can also be identified (13).

Analysis of data with this model-free method may help the experimentalist to develop an appropriate mathematical model, by indicating which aspects of the data are invariant and reproducible between curves. However, we still have an empirical approach, with the major purpose of providing a description of the fitted function. In contrast, an appropriately chosen mathematical model may provide parameters with a direct physicochemical interpretation, provide clues to mechanisms, or permit extrapolation beyond the range of the data. Thus, model-free analysis may be less informative, and will not substitute for properly implemented mathematical modeling.

By extending some of the advantages of mathematical modeling to families of curves of empirically defined, arbitrary shape, this method imposes some of the complexity and terminology of modeling. It requires some familiarity with basic principles of statistics, and necessitates the use of the computer. However, we have provided a practical implementation of the algorithm in a program, designated FLEXIFIT, for popular microcomputers. The program is easily used interactively, and provides graphic displays of data. Computing times are acceptable, and tend to be slow only for large data sets. The continuous development of computer hardware will allow for even faster, easier-to-use implementation.

Model-free analysis of data could be, as with any analytical or statistical technique, abused or misused. The user must select appropriate constraints, or the parameters may be ambiguous and nonidentifiable. One might inappropriately use an excessive number of scaling parameters to obscure large degrees of random or systematic variation. To avoid such problems, it is desirable to report the individual scaling factors, or at least their mean and standard deviations, and to show the results both in the original coordinates and after scaling to achieve optimal superimposition.

Further extensions and developments of this approach involve the development of a library of shape templates corresponding to defined mathematical models (e.g., parabola, hyperbola, logistic, exponential, power function, etc.). This might permit a selection of a mathe-

mathematical model for further analysis. In application to routine assays, a common reference shape can be extracted from 10 or 20 most recent assays of the same type. This shape template can subsequently be held constant, and every new assay curve can be tested for compatibility with the previously established template.

In summary, we have presented a method for model-free analysis of families of curves, based on weighted, four-parameter constrained smoothing splines. This approach can expand the role of computerized curve fitting to cases where a mathematical model is not available or not practical. By so doing, it could introduce more scientists to the principles of statistical analysis of data, provide an improved level of objectivity, and stimulate the search for appropriate mathematical representations of the underlying processes. EJ

Ms. Kim Chen developed the version of the program in FORTRAN 77. We wish to thank many colleagues who provided a constructive critique of preliminary versions of this manuscript, especially J. Ashwell, I. D. J. Bross, K. Dixon, D. Finney, S. Garattini, R. Jernigan, D. Lichtstein, J. E. A. McIntosh, J. E. Rall, M. Rocchetti, E. Rovati, S. Sabol, A. N. Schächter, R. H. Schwartz, S. Schwarz, L. Sheiner, and R. Shrager. We are grateful to several individuals who provided data used for testing the program: Drs. P. Ehrlich, W. Moyle, H. Xu, D. Lichtstein, J. Veldhuis, M. Maggi, B. Cox, C. Frash, M. Zweig, J. Foy, N. Beaudry, and R. Tacey. Ms. Jeanne Barnett and Ms. Mary Keffer provided excellent secretarial assistance.

REFERENCES

1. MOTULSKY, H. J.; RANSNAS, L. A. Fitting curves to data using nonlinear regression: a practical and nonmathematical review. *FASEB J.* 1: 365-374; 1987.
2. LAWTON, W. H.; SYLVESTRE, E. A.; MAGGIO, M. S. Self modeling nonlinear regression. *Technometrics* 14: 513-532; 1972.
3. REEVES, R. L.; KAIN, R. S.; MAGGIO, M. S.; SYLVESTRE, E. A.; LAWTON, W. H. Analysis of the visual spectrum of methyl orange in solvents and in hydrophobic building sites. *Can. J. Chem.* 51: 628-635; 1973.
4. REEVES, R. L.; MAGGIO, M. S.; COSTA, L. F. Importance of solvent cohesion and structure in solvent effects on binding site probes. *J. Am. Chem. Soc.* 96: 5917-5925; 1974.
5. GRAHAM, B. V. Wavelength discrimination derived from color naming. *Vision Res.* 16: 559-562; 1976.
6. LEVINE, H. D.; ROSEN, A. L.; DEWOSKIN, R.; MOSS, G. S. Application of self-modeling nonlinear regression to ventricular pressure data. *Comput. Biomed. Res.* 10: 363-372; 1977.
7. ARMSTRONG, P. W.; WATTS, D. G.; HAMILTON, D. C.; CHIONG, M. A.; PARKER, J. O. Quantification of myocardial infarction: template model for serial creatine kinase analysis. *Circulation* 60: 856-865; 1979.
8. REINSCH, C. H. Smoothing by spline functions. *Numerische Math.* 10: 177-183; 1967.
9. RODBARD, D.; LENOX, R. H.; WRAY, H. L.; RAMSETH, D. Statistical characterization of the random errors in the radioimmunoassay dose-response variable. *Clin. Chem.* 22: 350-358; 1976.
10. DELEAN, A.; MUNSON, P. J.; RODBARD, D. Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves. *Am. J. Physiol.* 235: E97-E102; 1978.
11. MUNSON, P. J.; RODBARD, D. LIGAND: a versatile computerized approach for characterization of ligand-binding systems. *Anal. Biochem.* 107: 220-239; 1980.
12. RODBARD, D.; MUNSON, P. J.; DELEAN, A. Improved curve fitting, parallelism testing, characterization of sensitivity and specificity, validation, and optimization for radioligand assays. *Radioimmunoassay and related procedures in medicine*, vol. 1. Vienna: International Atomic Energy Agency; 1978: 469-504.
13. GUARDABASSO, V.; RODBARD, D.; MUNSON, P. J. A model-free approach to estimation of relative potency in dose-response curve analysis. *Am. J. Physiol.* 252: E357-E364; 1987.
14. PORIER, D. J. Piecewise regression using cubic splines. *J. Am. Stat. Assoc.* 68: 515-524; 1973.
15. WOLD, S. Spline function in data analysis. *Technometrics* 16: 1-11; 1984.
16. MANNERVIK, B. Regression analysis, experimental error, and statistical criteria in the design and analysis of experiments for discrimination between rival kinetic models. *Methods Enzymol.* 87: 370-390; 1982.
17. MAGAR, M. E. *Data analysis in biochemistry and biophysics*. New York: Academic; 1972.
18. CLEVELAND, U. S. Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.* 74: 829-836; 1979.
19. DRAPER, N. R.; SMITH, H. *Applied regression analysis*. New York: Wiley; 1966.
20. MARQUARDT, D. W. An algorithm for least squares estimation of non-linear parameters. *J. Soc. Ind. Appl. Math.* 2: 431-441; 1963.
21. BENNETT, C. A.; FRANKLIN, N. L. *Statistical analysis in chemistry and the chemical industry*. New York: Wiley; 1954.
22. EFRON, B. *The jackknife, the bootstrap and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics; 1982.
23. DEBOOR, C. A. *Practical guide to splines*. New York: Springer-Verlag; 1978.

APPENDIX

A full discussion of properties of splines can be found in standard references such as deBoor (23). We summarize briefly the definition of a spline. Given the sequence of knots $x_1 < x_2 < \dots < x_n$, and a particular value of x , where $x_i \leq x < x_{i+1}$, for some $i = 1, \dots, n - 1$, the spline function is a cubic polynomial

$$f_s(x) = p_i + q_i(x - x_i) + r_i(x - x_i)^2 + s_i(x - x_i)^3$$

The set of coefficients (p_i, q_i, r_i, s_i) , $i = 1, \dots, n - 1$, for the spline polynomials must obey the following conditions (setting $h_i = x_{i+1} - x_i$):

1) Continuity of the spline:

$$p_i + q_i h_i + r_i h_i^2 + s_i h_i^3 = p_{i+1} \quad i = 1, \dots, n - 2$$

2) Continuity of the first derivative:

$$q_i + 2r_i h_i + 3s_i h_i^2 = q_{i+1} \quad i = 1, \dots, n - 2$$

3) Continuity of the second derivative:

$$2r_i + 6s_i h_i = 2r_{i+1} \quad i = 1, \dots, n - 2$$

4) End conditions:

$$r_1 = 0 \text{ and } r_{n-1} + 6s_{n-1}h_{n-1} = 0$$

Given a set of data (x_i, y_i) $i = 1, \dots, n$, a smoothing spline minimizes the integral $\int_{x_1}^{x_n} [f_S''(x)]^2 dx$, subject to the condition $\sum_i [y_i - f_S(x_i)]^2 \leq S$ for a specified smoothing factor S . The particular values for the parameters (p, q, r, s) may be calculated with the algorithm described by Reinsch (8). In our algorithm the factor S is adjusted to the smallest possible value (i.e., least degree of smoothing) for which the user-specified constraints of monotonicity and/or number of inflection points are satisfied.

A particular member of a family of curves of similar shape may be represented as a linear shift and scale, in both horizontal and vertical directions, around an underlying shape function f_S , in this case a spline function. For curve j , the model is

$$y = (a_j - d_j)f_S[b_j(x - c_j)] + d_j$$

Here, the values of a and d may be regarded as defined landmark values in the y measurement scale, e.g., baseline and maximal response levels. The parameter c is horizontal shift; b is a horizontal stretch or scaling factor.

The FLEXIFIT algorithm for adjusting up to four parameters (a_j, b_j, c_j, d_j) for each curve together with the parameters of the spline function is as follows:

1) Specify initial estimates for (a_j, b_j, c_j, d_j) for $j = 2, \dots, m$, the number of curves in the family. The values for (a_1, b_1, c_1, d_1) can be fixed at $(1, 1, 0, 0)$.

2) Shift and scale the data from all the curves,

for curve 1: $x'_i \leftarrow x_i$

$$y'_i \leftarrow y_i$$

for curve 2: $x'_i \leftarrow b_2(x_i - c_2)$

$$y'_i \leftarrow (y_i - d_2)/(a_2 - d_2)$$

for curve j : $x'_i \leftarrow b_j(x_i - c_j)$

$$y'_i \leftarrow (y_i - d_j)/(a_j - d_j)$$

3) Fit the smoothing spline f_S to the scaled and shifted data (x'_i, y'_i) , $i = 1, \dots, n$, adjusting the smoothing factor S to satisfy user-specified constraints concerning monotonicity and number of inflection points.

4) Now fix the parameters (p_i, q_i, r_i, s_i) of the smoothing spline f_S . Using a Levenberg-Marquardt algorithm (20), adjust the parameters (a_j, b_j, c_j, d_j) of the full model

$$y = (a_j - d_j)f_S[b_j(x - c_j)]$$

to the data for each of the $j = 1, \dots, m$ curves, so that the sum-of-squares of residuals is minimized.

5) Iterate steps 2, 3, and 4 until convergence is obtained (the weighted sum-of-squares of residuals changes less than 0.01%).

6) Calculate summary statistics regarding the fit, including the parameters (a, b, c, d) , their standard errors, confidence intervals, root-mean-square error (RMS), number of sign runs of residuals, mean square successive difference (MSSD) test applied to the residuals, appropriate ratios or differences of parameters [e.g., log relative potency = $(c_2 - c_1)$ when x is log dose].

Constraints involving the various parameters should be employed to avoid over-parameterization. Selected groups of parameters may be required to share a value (e.g., $a_1 = a_2$) or to remain at predetermined constant values (e.g., $b_2 = 1$). Judicious use of this feature allows for testing relevant hypotheses about these parameters by comparing the residual sum-of-squares (and other criteria for goodness-of-fit) between runs with different constraints.

Several alternative parameterizations are possible. The parameter c_i is used here to correspond to $\log c_i$ in ALLFIT (10). It may also be desirable to set $K_i = a_i - d_i$ (12). Each of these parameterizations has its own advantages and may be useful in selected applications.